68015

CRFE

Please rush. Thanks Chris

-----Original Message-----
From: Sorbello, Eleanor
Sent: Tuesday, June 04, 2002 1:28 PM
To: Chan, Christina
Cc: Sorbello, Eleanor
Subject: Rush seq search 09/515,369B

Christina,
Can you please reqest a rush seq. search of (1) **nucleotide SEQ. ID. NO: 1** and (2) any matches from **T at position -2241 to the C at posn. 0 of Seq. Id NO: 1**

Eleanor Sorbello
Art Unit 1632 ✓
CM1 12D14 ✓
(703) 308-6043

12E12

Numbering does not correspond to our database system of numbering.

In our files the first 10 positions are: taatacgact
(1 ... 10)

There is no zero position and there are no negative numbers.

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

TYPE OF SEARCH:	VENDOR/COST(where applic.)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: __6/3__
Date Completed: __6/6__
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

OM nucleic - nucleic search, using sw model

```
                    GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: June 6, 2002, 08:23:33 ; Search time 4143.56 Seconds
(without alignments)
11545.160 Million cell updates/sec

Title: US-09-515-369B-1
Perfect score: 2286
Sequence: 1 taatacgactcactaggg.........tgacttccacggctgggacg 2286

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046326B293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0:
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
  Maximum Match 100%
  Listing first 45 summaries

```
Database : GenEmbl:*
  1: gb_ba:*
  2: gb_htg:*
  3: gb_in:*
  4: gb_om:*
  5: gb_ov:*
  6: gb_pat:*
  7: gb_ph:*
  8: gb_pl:*
  9: gb_pr:*
 10: gb_ro:*
 11: gb_sts:*
 12: gb_sy:*
 13: gb_un:*
 14: gb_vi:*
 15: em_ba:*
 16: em_fun:*
 17: em_hum:*
 18: em_in:*
 19: em_mu:*
 20: em_om:*
 21: em_or:*
 22: em_ov:*
 23: em_pat:*
 24: em_ph:*
 25: em_pl:*
 26: em_ro:*
 27: em_sts:*
 28: em_un:*
 29: em_vi:*
 30: em_htg_hum:*
 31: em_htg_inv:*
 32: em_htg_other:*
 33: em_htgo_inv:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%	Length	DB	ID	ID	Description
1	2255	98.6	2256	9	AF217405	AF217405	Homo sapi
2	2208.6	96.6	193317	2	AC098935	AC098935	Homo sapi
3 c	2208.6	96.6	229302	2	AC023534	AC023534	Homo sapi
4	2151.4	94.1	180048	2	AC068122	AC068122	Homo sapi
5	1872.2	81.9	11114	9	AY062931	AY062931	Homo sapi
6	1365.2	59.7	7025	9	AF276916	AF276916	Homo sapi
7	91.8	4.0	29034	9	AJ243947	HSA243947	Homo sapi
8 c	91.2	4.0	174210	2	AC108684	AC108684	Homo sapi
9 c	90.8	4.0	193126	2	AC002090	AC002090	Genomic s
10 c	90	4.0	177957	9	AL139275	AL139275	Human DNA
11	90	3.9		9	AC025298	AC025298	Homo sapi
12 c	89.2	3.9	178477	9	AC022801	AC022801	Homo sapi
13	89.2	3.9	192695	2	AC098645	AC098645	Papio cyn
14	89.2	3.9	181009	2	AC011060	AC011060	Homo sapi
15	88.4	3.9	183483	2	AL161628	AL161628	Human DNA
16	88.2	3.9	194590	2	AL445623	HS315114	Homo sapi
17	88.2	3.9	168486	9	AL163872	CNS05TBN	Human chr
18	88.2	3.9	149973	2	AC023992	AC023992	Homo sapi
19 c	88.2	3.9	159905	9	AC005566	AC005566	Homo sapi
20 c	87.4	3.8	163242	9	AC068438	AC068438	Homo sapi
21 c	87	3.8	181134	2	AC074343	AC074343	Homo sapi
22	87	3.8	62497	2	AC103865	AC103865	Homo sapi
23 c	86	3.8	178756	9	AC104420	AC104420	Homo sapi
24 c	85.4	3.8	116516	2	AC109809	AC109809	Homo sapi
25 c	85.4	3.7	86025	9	AL158175	AL158175	Human DNA
26 c	85.4	3.7	119361	2	AC092771	AC092771	Homo sapi
27 c	85.4	3.7	142192	2	AC092713	AC092713	Homo sapi
28 c	85.4	3.7	158341	2	AC021733	AC021733	Homo sapi
29 c	85	3.7	160671	2	AC090451	AC090451	Homo sapi
30 c	85	3.7	162625	9	AC006083	AC006083	Homo sapi
31 c	85	3.7	127936	2	AC022099	AC022099	Homo sapi
32 c	84.6	3.7	143536	2	AC020846	AC020846	Mus muscu
33 c	84.6	3.7	172830	9	AL137072	AL137072	Human DNA
34 c	84.6	3.7	59319	2	AC103842	AC103842	Homo sapi
35 c	84.6	3.7	114467	9	AP000462	AP000462	Homo sapi
36 c	84.6	3.7	154236	9	AC051650	AC051650	Homo sapi
37	84.4	3.7	157322	9	AC012103	AC012103	Homo sapi
38 c	84.4	3.7	157399	9	AP003167	AP003167	Homo sapi
39	84.4	3.7	170965	9	AB017654	AB017654	Homo sapi
40 c	84.4	3.7	144714	9	AC020891	AC020891	Homo sapi
41	84	3.7	145886	9	AP003438	AP003438	Homo sapi
42	83.8	3.7	175830	2	AC090547	AC090547	Homo sapi
43 c	84	3.7	188764	2	AC021238	AC021238	Homo sapi
44	84	3.7	170893	9	AC027537	AC027537	Homo sapi
45 c	83.8	3.7	37321	9	AC005498	AC005498	Homo sapi

ALIGNMENTS

```
RESULT 1
AF217405

LOCUS       AF217405                2256 bp    DNA     linear   PRI 19-SEP-2000
DEFINITION  Homo sapiens suppression of tumorigenicity 16 protein (ST16) gene,
            promoter and partial sequence.
ACCESSION   AF217405
VERSION     AF217405.1  GI:10185703
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2256)
AUTHORS     Madireddi,M.T., Dent,P. and Fisher,P.B.
TITLE       AP-1 and C/EBP transcription factors contribute to mda-7 gene
            promoter activity during human melanoma differentiation
JOURNAL     J. Cell. Physiol. 185 (1), 36-46 (2000)
MEDLINE     20400556
PUBMED      10942517
REFERENCE   2  (bases 1 to 2256)
AUTHORS     Madireddi,M.T. and Fisher,P.B.
TITLE       Direct Submission
JOURNAL     Submitted (17-DEC-1999) Urology, Columbia University, 630 West,
```

FEATURES
source   168th Street, BB 1505, New York, NY 10032, USA
            Location/Qualifiers
            1..2256
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="placenta"
promoter   1..2208
            /gene="ST16"
gene   1..>2256
            /gene="ST16"
            /note="mda-7"
mRNA   2209..>2256
            /gene="ST16"
BASE COUNT 539 a   584 c   503 g   630 t
ORIGIN
            /product="suppression of tumorigenicity 16 protein"

Query Match 98.6%;   Score 2255;   DB 9;   Length 2256;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 2255;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   32 tcacctttgaaccaggtcgctggctgcctccaaagcttgtactcataactagattctcaa 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TCACCTTTGAACCAGGTCGCTGGCTGCCTCCAAAGCTTGTACTCATAACTAGATTCTCAA 60

Qy   92 ctgatgtgggccaaggtcctagttctctcctgaccttcctcctctgaagtaataatgc 151
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTGATGTGGGCCAAGGTCCTAGTTCCTAGGTTCCTCCTCTGACCTTCCTCCTGAAGTAATAATGC 120

Qy   152 tatgataagctcatcggaggctgaggcccaggccacatgtttgcctgaactatccatgtta 211
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TATGATAAGCTCATCGGAGGCCCAGGCCACAGGCCACATGTTTGCCTGAACTATCCATGTTA 180

Qy   212 tatgattcctcctcagacagagtgagctactcacgatcccagtgtacccttgaggccag 271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TATGATTCCTTCCTTCCTCAGACAGAGTGAGCTACTCACGATCCCAGTGTACCCTGAGGCCAG 240

Qy   272 ccaaggtgtatccatgactcatgcctctgttccagcctgccctttaacagcatcatccca 331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCAAGGTGTATCCATGACTCATGCCTCTGTTCCAGCCTGCCCTTTAACAGCTCATCCCA 300

Qy   332 cctgtcctcccccgcctatctgcagacagtagtctcagggattcagctgcctgtggggg 391
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCTGCCTGCCCTCCCGCCTATCTGCAGACAGTAGTCTCGGGATTTCAGCTGCCCTGGGGG 360

Qy   392 ctcattttccctccagctcagcttctgtctctgttagctgcctcccactcacctattact 451
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CTCATTTTCCCTCCAGCTCAGCTTCCTGTCTCTGCTTAGCTGCCTCCCACTCACCTATTACT 420

Qy   452 ccagcactctgtcttctttttgtctcatcactgctcttgacatcttatctc 511
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCAGCACTCTCTGTCTTCTTTTTGTCTCTCCTTCATCACTGCCCTGACATCTTTATCTC 480

Qy   512 atagtagttaggggttcttggtaatgccctaaatccacatggtgggaagggggag 571
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATAGTAGTTAGTTAGGGGTTCTTGGTAATGCCCTAAATCCACATGGTGGGAAGGGGGAG 540

Qy   572 tgggggaagagagtggcgctggggggtgaagactcgggccct 631
           ||||||||||||||||||||||||||||||||||||||||||
Db   541 TGGGGGTTCCAGAGTGGCGCTGGGGGGTGAGGGTAAGACTCGGGCCCT 600

Qy   632 ccaggaacaaaggattcaggctggtgcagctatagccaagcagacagctggccaggggat 691
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CCAGGAACAAAGGATTCAGGCTGGTGGCAGCTATAGCCAAGCAGACAGCTGGCCAGGGAT 660

Qy   692 tgcaaaggagtatttgtttgcttaagaaaataaacaacactgagtatgagatgggagga 751
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TGCAAAGGAGTATTTGTTTGCTTAAGAAAATAAACAACACTGAGTATGAGATGAGGGGA 720

Qy   752 gggggtgtggtggtgccagagagattgggaagagtctgccaaggtgttcttactcactct 811
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GGGGGTGTTGGTGGTGCCAGAGAGAGTTGGGAAGAGTCTGCCAAGGTGTGTTCTTACTCACTCT 780

Qy   812 cctctttttcttcatctccactgagctggaggcagttacccgtgcccccagtgtcacattc 871
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CCTCTTTTCTTCTTCATCTCCACTGAGCTGGGAGGCAGTTATCCCGTGCCCCCACGTCACATTC 840

Qy   872 ctactcccgttccatgcctggaccagtgtgggcaaactcttcctgtaagaaccaga 931
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CTACTCCCGTTCCCATGCCTGGACCCAGGTTGGGCAAACTCTTCCTGTAAAGAACCAGA 900

Qy   932 caggaactattttaggctctgtgtgccatatggtctcagtcacaactactcatcctctgcc 991
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CAGGAACTATTTTAGGCTCTGTGTGCCATATGGTCTCAGTCACAACTACTCATCTCTGCC 960

Qy   992 tctgtagcacgaaagcaattagcaacaatatgtcaacaaacatatgtgaccccatgaaaa 1051
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TCTGTAGCACGAAAGCAATTAGCAACAATATGTCAACAAACATATGTGACCCCATGAAAA 1020

Qy   1052 ctttattattatggatacggaacctgaaaataatgtctttccttctttttgatttttcccca 1111
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTTTATTTATTATGGATACGGATACCTGAAAATAATGTCTTTCCTTTTCTTTTGATTTTTTTCCCCA 1080

Qy   1112 atcattaaaaaacgtaaaaactactcttaggtcgcaaggttaagccattctcagcttagc 1171
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ATCATTAAAAAACGTAAAAACTACTTCTTAGGTCGCAAGGTTAAGCCATTCTCAGCTTAGC 1140

Qy   1172 agtggcaggctggattggcttgtgacctacagttggccaatcctgattcccaaaatgt 1231
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AGTGGCAGGCTGGATTTGGCTTGTGACCTACAGTTGGCCAATCCCTGATTCCCAAAATGT 1200

Qy   1232 attcctcaggtatgtgggcaaatacttatgggaagtgctggattaaacagagttaagaag 1291
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATTCCTCAGGTATGTGGGCAAATACTTATGGGAAGTGCTGGATTAAACAGAGTTAAGAAG 1260

Qy   1292 catcagacatttccaggacgggtagcacatgccaggtctctaactgaccttcattgga 1351
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CATCAGACATTCCAGGACGGGCTAGCACATGCCAGGGCTCTCTAACTGACCTCATTGA 1320

Qy   1352 ttcatctgtttcatggagatctgcaagacaagaatcctcaaccctagagtctgagga 1411
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCATCTGTTTCATGGAGGATCTGCAAGACAAGAATTCCTCAAAACCTAGAGTCTGAGGA 1380

Qy   1412 ctgtgcttgggaaacactgctctgctctgagtccactgggcacatggtagaatctag 1471
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CTGTGCTTGGGAAAACACTGCTCTGCTCTGAGTCTGGGCACATGGGCACATGGTAGAATCTAG 1440

Qy   1472 agctgagtgcctgctagctggagataggtcagagctcttgactgcctggccagttctg 1531
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGCTGAGTGCCTTGCTAGCTGGAGATAGGTCAGAGCTCTTGACTGCCCTGGCAGTCTTG 1500

Qy   1532 acacatccagctgtctgtctgtcccctgagtggtcagagccacacaggccaagactagccc 1591
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ACACATCCAGCTGTCTGTGTCCCTGAGTGGTTCAGAGCCACACAGGCCAAGACTAGCCC 1560

Qy   1592 accagacaccaggcctccagctttctggccttgtcatggctgtacatttcctattctt 1651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACCAGACACCAGGCCTCCAGCTTTCTGGGCTTGTGCCATGCGTACATTTCCTTATTCTT 1620

Qy   1652 cctggttccagaacctaaggagaggcacatttggtgagtgattataacctaggac 1711
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CCTGGTTCCAGAACCTAAGGAGAGGCACATTTTGGTTGAGTGATTATAACCCTAGGAC 1680

Qy   1712 catgggtagctgcatgtgcagcagtgccctgatgcctgatgattaaagga 1771
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CATGGGTAGCTGCATGTGCACGTGCCCGCCCTGATGCCCTGATGATTAAAGGA 1740

Qy   1772 gaggtacttacaggttattctcgctgtggactactgtcccagcacatgaatagggcatca 1831
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAGGTACTTACAGGTTATTTCTTCGCTGTGGACTACTGTCCCAGCACATGAATAGGGCCATCA 1800

Qy   1832 tattgaattatttgacaggaggagggactggtgtatgctgcacagtaataatgtattta 1891
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TTATTGAATTATTTGACAGGAGGAGGGACTGGTGTATCCTGCACAGTAATAATGTATTTA 1860

```
Qy    1892  catgtgtacagagtttaccaagcacctctgtgttgttttgccttgttattacacttg  1951
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861  CATGTGTACAGAGTTTACCAAGCACCTCGTGTTGTTTTGCCTTTGTTATTACACTTG  1920

Qy    1952  ggacaaattttaaaattatacgcagagactgcagcgcagagaagctaagagacttg  2011
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921  GGACAAATTTTAAAATTTATACATGCAGAGACTGCAGCGCAGAGAAGCTAAGAGACTTG  1980

Qy    2012  ccctgccacacagccagtggtagagcctgaactcaaaccaggtctcatctcaccctca  2071
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981  CCCCTGCCCACACAGCCAGTGGTAGAGCCTGAACTCAAACCCAGGTCTCATCTCACCTCA  2040

Qy    2072  gggctgcttccccatcgtcgtcttattgtcctaaagtgatgggtgactaggcaatgaagt  2131
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041  GGGGCTGCTTCCCCATCGTGTATTGTCCTTAAAGTTGATGGGTGACTAGGCAATGAAGT  2100

Qy    2132  aattctctaggaaagcatgacaattccctttcctcccactccctctttttcctccaccc  2191
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101  AATTCTCTAGGAAAGCATGACAATTTCCCTTTCCTCCCACCTCCCTCTTTTTCCTCCACCC  2160

Qy    2192  ctccccatcagcccccatatatatgccaaaatctccacaaagccttgcttgctgcaaa  2251
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161  CTCCCCATCAGCCCCCATATATATGCCCAAAATCTCCACAAAGCCTTGCTTGCTGCAAA  2220

Qy    2252  cctttacttctgaaatgacttccacggctgggacg  2286
            ||||||||||||||||||||||||||||||
Db    2221  CCTTTACTTCTGAAATGACTTCCACGGCTGGGACG  2255
```

```
RESULT  2
AC098935
LOCUS        AC098935           193317 bp    DNA         linear   HTG 06-NOV-2001
DEFINITION   Homo sapiens chromosome 1 clone RP11-564A8, WORKING DRAFT SEQUENCE,
             2 unordered pieces.
ACCESSION    AC098935 AL359089
VERSION      AC098935.1  GI:16751905
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 193317)
AUTHORS      Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 193317)
AUTHORS      Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (06-NOV-2001) Genome Center, University of Washington,
             Box 352145, Seattle, WA 98195, USA
COMMENT      On Nov 6, 2001 this sequence version replaced gi:10277966.
             ----------------------- Genome Center
             Center: University of Washington Genome Center
             Center Code: UWGC
             Web site: http://www.genome.washington.edu
             Contact: uwgchtgs@u.washington.edu
             Drafting Center: SC
             ----------------------- Project Information
             Center project name: chr-1
             Center clone name: RP11-564A8 (sc0742)
             ----------------------- Summary Statistics
             Sequencing vector: plasmid; 39% of reads
             Sequencing vector: plasmid; L08752; 61% of reads
             Chemistry: Dye-terminator ET; 89% of reads
             Chemistry: Dye-terminator Big Dye; 11% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 192768 bases at least Q40
             Consensus quality: 193102 bases at least Q30
             Consensus quality: 193186 bases at least Q20
             Insert size: 193217; sum-of-contigs
             Quality coverage: 10.0x in Q20 bases; sum-of-contigs
             ----------------------
             * NOTE: This is a 'working draft' sequence. It currently
```

```
             * consists of 2 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *          1    56155: contig of 56155 bp in length
             *      56156    56255: gap of unknown length
             *      56256   193317: contig of 137062 bp in length.
FEATURES             Location/Qualifiers
     source          1..193317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-564A8"
                     /clone_lib="RPCI human BAC library 11"
                     1..56155
     misc_feature    /note="assembly_name:Contig81"
                     56256..193317
     misc_feature    /note="assembly_name:Contig82"
BASE COUNT    53599 a  42586 c  43895 g  53137 t    100 others
ORIGIN
```

```
Query Match                 96.6%;   Score 2208.6;   DB 2;   Length 193317;
Best Local Similarity        98.8%;    Pred. No. 0;
Matches 2253;  Conservative   0;  Mismatches   9;  Indels  19;  Gaps   2;
```

```
Qy      25  gactcgatcacctttgaaccaggtctgcctgcctccaaagctcgtactcataactaga    84
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103494  GGCTCGATCACCTTTGAACCCAGGTCTGCCTGCCTCCAAAGCTTGTACTCATAACTAGA  103553

Qy      85  ttctcaactgatgttggccaaggttctctcctctccttagttctcctctgaacctcctgaagta   144
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103554  TTCTCAACTGATGTTGGGCCAAGGTTCCTCAGGTTCTCTCCTTGAACCTTCCTTCTTCTGAAGTA  103613

Qy     145  ataatgctatgataagctcatcgagagctgagagggccaggccagcacatgtttgctgaactatc   204
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103614  ATAATGCTATGATAAGCTCATCGGAGCTGAGGCTGAGGCCCAGGCACACATGTTTGCCTGAACTATC  103673

Qy     205  catgttatatgatttcctcagacagagtgagctactcacgatcccaggtgtacccctg   264
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103674  CATGTTATATGATTCCTTCCTCAGACAGAGTGAGCTACTCACGATCCCAGGTGTACCCCTG  103733

Qy     265  aggccagccaaggtgtatccatgacctcatgcctcgttccagcctgcccttaacagct   324
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103734  AGGCCAGCCAAGGTGTATCCATGACCTCATGCCTCGTTCCAGCCTGCCCTTAACAGCT  103793

Qy     325  catcccacctgcctgcctccccgccatccgcctccccgccccctcatctgcagacagtagtctgagatttcagctgcc   384
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103794  CATCCCACCTGCCTGCCTCCCCGCCTATCTGCAGACAGTAGTCTAGGATTTCAGCTGCC  103853

Qy     385  ctggggctcatttcccctccagctcctgctttagctgtctcctgccccactcacc   444
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103854  CTGGGGCTCATTTCCCTCCAGCTCCTTCAGCTTCTTAGTGTCCTGCCTCCCACTCACC  103913

Qy     445  tattactccagcactctcactgtcttctttttgtctcatcactgctcctttgacatct   504
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103914  TATTACTCCAGCACTCTCACTGTCTTCTTTTTGTTCTGTCTCATCACTGCTCCTTGACATCT  103973

Qy     505  ttatccatagtagttagtggggtctggtaatgccctaaatccacatgtgggaag   564
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103974  TTATCCATAGTAGTTAGTGGGGTTCTTGGTAATGCCCTAAATCCACATGGTGGGAAG  104033

Qy     565  ggggagtggggaagaggtgcggctgtgggggctgtggctgtccacttctggaggtaagactc   624
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104034  GGGGAGTGGGGAAGAGAGTGCGGCTGTGGGGCTGTGCCTACTTCTGGAGGGTAAGACTC  104093

Qy     625  gggccctccaggaacaaagattcaggctggtggcagctatagccaagcagagagactgctggc   684
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104094  GGGCCCTCCAGGAACCAAAGGATTCAGGCTGGTGGCAGCTATAGCCAAGCCAAGCAGAGACTGCTGGC  104153
```

Left column:

```
Qy    685 cagggattgcaaaggagtatttgttgctgagaaataaacaacactgagtagtatgagat 744
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104154 CAGGGATTGCAAAGGAGTATTTGTTGCTTAAGAAATAAACAACACTGAGTATGAGAT 104213

Qy    745 ggagggaggggggtgttggtgccagagagattgggaagagtctgccaaggtgtgttctac 804
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104214 GGAGGGAGGGGGGTGTTGGTGCCAGAGAGATTGGGAAGAGTCTGCCAAGGGTGTGTTCTAC 104273

Qy    805 tcactctcctctttctttcttcttcatctccactgactggaggcagttatcctgtccccacgt 864
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104274 TCACTCTCCTCTTTCTTTCTTCATCTCCACTGACTGGAGGCAGTTATCCTGTCCCCACGT 104333

Qy    865 cacattcctactccgttccatgcctggcaccaggttgggcaaactcttcctgtaaag 924
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104334 CACATTCCTACTCCCGTTCCATGCCTGGCACCAGGTTGGGCAAACTCTTCCTGTAAAG 104393

Qy    925 aaccagcaggaactattttaggctgtcgtgccatatggttctcagtcacaactactcat 984
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104394 AACCAGACAGGAACTATTTTAGGCTGTCGTGTGCCATATGGTTCTCAGTCACAACTACTCAT 104453

Qy    985 cctgccctgtagcacgaaagcaattagcaacaatatgtcaacaaacatatgtgaccccc 1044
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104454 CTCTGCCCTGTAGCACGAAAGCAATTAGCAACAATAGTGTCAACAAACATATGTGACCCCC 104513

Qy   1045 atgaaacttatttattatgatacggaaacctgaaaataatgtctttctttttgattttt 1104
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db 104514 ATGAAAACTTTATTTATTATGATACGGAAAATAATGTCTTTCTTTTTGATTTT 104573

Qy   1105 ttcccaatcattaaaaaactaactcttaggtcgcaaggttaagccattctca 1164
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104574 TTCCCAATCATTAAAAAACGTAAAAACTACTCTTAGGTCGCAAGGTTAAGCCATTCTCA 104633

Qy   1165 gcttagcagtggcaggctggatttggcttgtgacctacagttggccaatccctgattccc 1224
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104634 GCTTAGCAGTGGCAGGCTGGATTTGGCTTGTGACCTACAGTTGGCCAATCCCTGATTCCC 104693

Qy   1225 aaaatgtattcctcaggagatgtgggcaaatacttatgggaagtgctggattaaaacagagt 1284
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104694 AAAATGTATTCCTCAGGAGATGTGGGCAAAATACTTATGGGAAATGCTGGATTAAAACAGAGT 104753

Qy   1285 taagaagcatcagacatttccaggacgggctagcacatgcacgcagggctctcaactgacct 1344
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104754 TAAGAAGCATCAGACATTTCAGGACGGGCTAGCACATGCACGCAGGGCTCTCAACTGACCT 104813

Qy   1345 catt-ggattcatctgtttcatggaggatcttgcaagacaagaattcctcaaacctagag 1403
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104814 CATTAGGATTCATCTGTTTCATGGAGGATCTTGCAAGACAAGAATTCCTCAAACCTAGAG 104873

Qy   1404 tctgaggactgtgcttgggaaacactgctctgtcttgatgccctcactgggcacatggta 1463
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104874 TCTGAGGACTGTGCTTGGGAAACACTGCTCTGTCTTGATGCCCTCACTGGGCACATGGTA 104933

Qy   1464 gaatctagagcgagctgccttgctagctgagatagggtcagagctcttgactgcctgg 1523
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104934 GAATCTAGAGCTGAGTGCCTTGCTAGCTGAGATAGGGTCAGAGCTCTTGACTGCCTGG 104993

Qy   1524 cagtcttgacatcagacattccaggctgtctgtccctgagtggttcagagccacagggcaag 1583
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104994 CAGTCTTGACACATCAGACATTCCAGGCTGTCTGTGTCCCTGAGTGGTTCACAGCCACCAGGGCAAG 105053

Qy   1584 actagcccaccagagcaccaggcctccaccaccagcttctgggctgttcatgcgtacatttcc 1643
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105054 ACTAGCCCACCACCAGAGCACCAGGCCTCCACCACCAGCTTCTGGGCTTGTCCATGCGTACATTTCC 105113

Qy   1644 ttattcctggtttccagaacctaaggagaggccacatttggttgagtgattataaac 1703
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105114 TTATTCCTGGTTTCCAGAACCTAAGGAGAGGCACAATTTGGTGAGTGATTATAAACC 105173

Qy   1704 ctaggagaccatgggagctgcatgtgcaatgtcaggaaaacactcctcaacttcctggccctgatgga 1763
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105174 CTAGGGACCATGGTAGCTGCATGTGCAATGTCAGGAAAAACACTCCTCAACTTCCTGGCCCTGATGGA 105233

Qy   1764 ttaaaggagaggtacttacagttattcttcgctgtggactactgtccccagcatgaata 1823
```

Right column:

```
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105234 TTAAAGGAGAGGTACTTACAGGTTATTTCTTCGGCTTCTCGCTGTGGACTACTACTGTCCCAGCCATGAATA 105293

Qy   1824 gggcatcattattgaattatttgacagaggagagactggtgtatgctgcacagtaataa 1883
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105294 GGGCATCATTATTGAATTATATTTGACAGGAAGGAGAGACTGGTGTGAATGCTGCACAGTAATAA 105353

Qy   1884 tgtattacatgtgtacagagtttaccaagcacctctgtgtgttttggccttgttat 1943
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105354 TGTATTTACATGTGTACAGAGTTTACCAAGCACCTCTGTGTGTTTTTGCCCTCTGTTTAT 105413

Qy   1944 tacactt----------gggacaaattttaaaattatacatgcagagact 1985
          ||||||            |  |||||||||||||||||||||||||||||||
Db 105414 TACACTTGCCCTGGGGGAGTGGGACGGGAGGTGGGCAAAAATTTTTAAAAATTTATACATGCAGAGACT 105473

Qy   1986 gcagcgcagagaaagctaagagacttgccctgcccctgccaccacagccagtggtaggccctgaac 2045
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105474 GCAGCGCAGAGAGCTAAGAGACTTGCCCTGCCCCTGCCACCACAGCCAGTGGTAGGCCTGAAC 105533

Qy   2046 tcaaaaccaggtctcatctcacctcaggggctgcttcccatcgctgtattgtccttaa 2105
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105534 TCAAACCCAGGTCTCATCTCACCTCAGGGGTGCTTTCCCCATCGCTGTATTGTCCTTAA 105593

Qy   2106 agtgatggtgactaggcaatgaagtaattctctaggaaagcatgaccaatttcccttttc 2165
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105594 AGTGATGGTGACTAGGCAATGAAGTAATTCTCTAGGAAAGCAATGACCAATTTCCCTTTTC 105653

Qy   2166 tccacctccctttttcctccaccctcccccatcagcccccatatatgcccaaatc 2225
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105654 TCCACCTCCCTCTTTTTCCTCCACCCTCCCCCATCAGCCCCCATATATGCCCAAATC 105713

Qy   2226 tccacaaagccttgctgctgccgtcgcaaacctttacttctgaaatgaatgacttccacgcgctggac 2285
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105714 TCCACAAAGCCTTGCTTGCTGCCGTCGCAAACCTTTACTTCTGAAATGACTTCCACGCGCTGGAC 105773

Qy   2286 g 2286
          |
Db 105774 G 105774

RESULT 3
AC023534/c
LOCUS       AC023534             229302 bp    DNA     linear   HTG 04-MAY-2000
DEFINITION  Homo sapiens clone RP11-462N18, WORKING DRAFT SEQUENCE, 41
            unordered pieces.
ACCESSION   AC023534
VERSION     AC023534.3  GI:7705140
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229302)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-462N18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229302)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,R., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,C.H., Mihova,T., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Norman,C.H., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
```

```
          Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
          Travers,M., Trigilio,J., Vassiliev,H., V111iev.R., Vo,A., Wilson,B.,
          Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
          Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 4, 2000 this sequence version replaced gi:7143452.
          All repeats were identified using RepeatMasker.
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
          -------- Project Information
             Center project name: L5244
             Center clone name: 462_N18
          -------- Summary Statistics
             Sequencing vector: M13; M77815; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.960731
             Consensus quality: 194690 bases at least Q40
             Consensus quality: 207622 bases at least Q30
             Consensus quality: 217061 bases at least Q20
             Insert size: 200000; agarose-fp
             Insert size: 225302; sum-of-contigs
             Quality coverage: 4.5 in Q20 bases; agarose-fp
             Quality coverage: 4.0 in Q20 bases; sum-of-contigs
          -----
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 41 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *       1  1108: contig of 1108 bp in length
        *    1109  1208: gap of     100 bp
        *    1209  2504: contig of 1296 bp in length
        *    2505  2604: gap of     100 bp
        *    2605  3688: contig of 1084 bp in length
        *    3689  3788: gap of     100 bp
        *    3789  4962: contig of 1174 bp in length
        *    4963  5062: gap of     100 bp
        *    5063  6375: contig of 1313 bp in length
        *    6376  6475: gap of     100 bp
        *    6476  7673: contig of 1198 bp in length
        *    7674  7773: gap of     100 bp
        *    7774  9425: contig of 1652 bp in length
        *    9426  9525: gap of     100 bp
        *    9526 10997: contig of 1472 bp in length
        *   10998 11097: gap of     100 bp
        *   11098 12740: contig of 1643 bp in length
        *   12741 12840: gap of     100 bp
        *   12841 14467: contig of 1627 bp in length
        *   14468 14567: gap of     100 bp
        *   14568 16980: contig of 2413 bp in length
        *   16981 17080: gap of     100 bp
        *   17081 19176: contig of 2096 bp in length
        *   19177 19276: gap of     100 bp
        *   19277 21882: contig of 2606 bp in length
        *   21883 21982: gap of     100 bp
        *   21983 23826: contig of 1844 bp in length
        *   23827 23926: gap of     100 bp
        *   23927 26561: contig of 2635 bp in length
        *   26562 26661: gap of     100 bp
        *   26662 28846: contig of 2185 bp in length
        *   28847 28946: gap of     100 bp
        *   28947 32495: contig of 3549 bp in length
```

```
        *   32496 32595: gap of     100 bp
        *   32596 34950: contig of 2355 bp in length
        *   34951 35050: gap of     100 bp
        *   35051 38193: contig of 3143 bp in length
        *   38194 38293: gap of     100 bp
        *   38294 41862: contig of 3569 bp in length
        *   41863 41962: gap of     100 bp
        *   41963 45899: contig of 3936 bp in length
        *   45899 45998: gap of     100 bp
        *   45999 50904: contig of 4906 bp in length
        *   50905 51004: gap of     100 bp
        *   51005 55346: contig of 4342 bp in length
        *   55347 55446: gap of     100 bp
        *   55447 61336: contig of 5890 bp in length
        *   61337 61436: gap of     100 bp
        *   61437 65806: contig of 4370 bp in length
        *   65807 65906: gap of     100 bp
        *   65907 71648: contig of 5742 bp in length
        *   71649 71748: gap of     100 bp
        *   71749 74886: contig of 3138 bp in length
        *   74887 74986: gap of     100 bp
        *   74987 79878: contig of 4892 bp in length
        *   79879 79978: gap of     100 bp
        *   79979 84280: contig of 4302 bp in length
        *   84281 84380: gap of     100 bp
        *   84381 89873: contig of 5493 bp in length
        *   89874 89973: gap of     100 bp
        *   89974 97168: contig of 7195 bp in length
        *   97169 97268: gap of     100 bp
        *   97269 101906: contig of 4638 bp in length
        *  101907 102006: gap of     100 bp
        *  102007 109439: contig of 7433 bp in length
        *  109440 109539: gap of     100 bp
        *  109540 117592: contig of 8053 bp in length
        *  117593 117692: gap of     100 bp
        *  117693 126037: contig of 8345 bp in length
        *  126038 126137: gap of     100 bp
        *  126138 135140: contig of 9003 bp in length
        *  135141 135240: gap of     100 bp
        *  135241 146384: contig of 11144 bp in length
        *  146385 146484: gap of     100 bp
        *  146485 160352: contig of 13868 bp in length
        *  160353 160452: gap of     100 bp
        *  160453 173458: contig of 13006 bp in length
        *  173459 173558: gap of     100 bp
        *  173559 191605: contig of 18047 bp in length
        *  191606 191705: gap of     100 bp
        *  191706 229302: contig of 37597 bp in length
FEATURES             Location/Qualifiers
     source          1..229302
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-462N18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1108
                     /note="assembly_fragment"
     misc_feature    1209..2504
                     /note="assembly_fragment"
     misc_feature    2605..3688
                     /note="assembly_fragment"
     misc_feature    3789..4962
                     /note="assembly_fragment"
     misc_feature    5063..6375
                     /note="assembly_fragment"
     misc_feature    6476..7673
                     /note="assembly_fragment"
     misc_feature    7774..9425
                     /note="assembly_fragment"
     misc_feature    9526..10997
                     /note="assembly_fragment"
     misc_feature    11098..12740
                     /note="assembly_fragment"
     misc_feature    12841..14467
```

```
misc_feature    14568. .16980    /note="assembly_fragment"
misc_feature    17081. .19176    /note="assembly_fragment"
misc_feature    19277. .21882    /note="assembly_fragment"
misc_feature    21983. .23826    /note="assembly_fragment"
misc_feature    23927. .26561    /note="assembly_fragment"
misc_feature    26662. .28846    /note="assembly_fragment"
misc_feature    28947. .32495    /note="assembly_fragment"
misc_feature    32596. .34950    /note="assembly_fragment"
misc_feature    35051. .38193    /note="assembly_fragment"
misc_feature    38294. .41862    /note="assembly_fragment"
misc_feature    41963. .45898    /note="assembly_fragment"

Query Match            96.6%;    Score 2208.6;    DB 2;    Length 229302;
Best Local Similarity  98.8%;    Pred. No. 0;
Matches 2253;    Conservative  0;    Mismatches  9;    Indels 19;    Gaps 2;

Qy      25 gactcgatcaccttttgaaccaggtctgcctgcctccaaagcttgtactcataactaga      84
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  158375 GGCTCGATCACCTTTTGAACCCAGGTCTGCCTGCCTCCAAAGCTTGTACTCATAACTAGA  158316

Qy      85 ttccaactgatgttgggccaaggtcctaggttctctcctccttcctctgaagta      144
           ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  158315 TTCTCAACTGATGTTGGGCCAAGGTCCTAGGTTCTCTAGGTTCCTTCCTTCTGAAGTA  158256

Qy     145 ataatgctatgataagcctcatcggagagctgagcgcccaggcacatgttgcctgaactatc      204
           |||||||||||||| ||||||||||||||||||||||||||||||||| ||||||||||||
Db  158255 ATAATGCTATGATAAGCCTCATCGGAGAGCTGAGCGCCCAGGCACAGTTTGCCTGAACTATC  158196

Qy     205 catgttatatgattccttcctcagacagagtgagctactcacgatccagtgtaccctg      264
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  158195 CATGTTATATGATTCCTTCCTCAGACAGAGTGAGCTACTCACGATCCCAGGTGTACCCTG  158136

Qy     265 aggccagccaaggtgtatccatgacctccatgcctctgttccagcctgccttaaacagct      324
           ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  158135 AGGCCAGCCAAGGTGTATCCATGACCTCCATGCCTCTGTTCCAGCCTGCCTTTAAACAGCT  158076

Qy     325 ctggggctcattttccctctcagctctgctttagctgtctcctgcctcccactcacc      384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  158075 CTGGGGCTCATTTTCCCTCTCAGCTCTGCTTTAGCTGTCTCCTGCCTCCCACTCACC  158016

Qy     385 tattactccagcactcaacctggtctctttctgtctcatcactgcctctttgacatct      444
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  158015 TATTACTCCAGCACTCAACCTGGTCTCTTTCTGTCTCATCACTGCCTCTTTGACATCT  157956

Qy     445 ttattactcatgatagtagttaggggttcttggttcttggtaatgccctaaatcacatggtgggaag      504
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157955 TATTACTCATGATAGTAGTTAGGGGTTCTTGGTTCTTGGTAATGCCCTAAATCACATGGTGGGAAG  157896

Qy     505 ttatggtcatgatagttagttagtgggtgagagagagtgcgctgtgggggtaagactc      564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157895 TTATGGTCATGATAGTTAGTTAGTGGGTGCGCTGTGGGGGTAAGACTC  157836

Qy     565 gggggagtggggggaagagagagtgcgctgtgggggctgtgtgcctacttctggagggtaagactc      624
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157835 GGGGGAGTGGGGGGAAGAGAGAGTGCGCTGTGGGGGCTAGGAGGGTAAGACTC  157776

Qy     625 gggccctccaggagaacaaaggattcaggctgcgctgttggcagctatagccaagcagactgctggc      684
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157775 GGGCCCTCCAGGAGAACAAAGGATTCAGGCTGCGGTGGCCATATAGCCAAGCAGACTGCTGGC  157716

Qy     685 caggggattgcaaaggagagtgtatttgttgcttaagaaaataaacaacactactgagatatgagat      744
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  157715 CAGGGATTGCAAAGGAGTATTTGTTGCTTAAGAAAATAAACAACACTGAGTATGAGAT  157656

Qy     745 ggaggggagggggtgttggtggtgcagagagattggaagagtctgccaaggtgtgttctac      804
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157655 GGAGGGAGGGGGTGTTGGTCCAGAGAGTTGGAAGAGTCTGCCAAGGTGTGTTCTAC  157596

Qy     805 tcactctcctttctcatctccactgagctgaggcagttatcctgtcccccacgt      864
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157595 TCACTCTCCTCTTTCATCTCCACTGAGCTGAGGCAGTTATCCTGTCCCCCACGT  157536

Qy     865 cacattcctactcccgtttccatgcctgacccaggttgggcaaactcttcctgtaaag      924
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157535 CACATTCCTACTCCCGTTTCCATGCCTGACCCAGGTTGGGCAAACTCTTCCTGTAAAG  157476

Qy     925 aaccagacaggaactatttaggctgtgtgccatatggtctcagtcacaactactcat      984
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157475 AACCAGACAGGAACTATTTAGGCTCTGTGTGCCATATGGTCTCAGTCACAACTACTCAT  157416

Qy     985 ctctgccttgtagcacgaaagcaattagcaacaatatgtcaacaaacatatgtgacccc      1044
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157415 CTCTGCCTCTGTAGCACGAAAGCAATTAGCAACAATATGTCAACAAACATATGTGACCCC  157356

Qy    1045 atgaaaacttattattattaggatacggaaacctgaaaataatgtctttccttttgatttt      1104
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157355 ATGAAAACTTTATTTATTTATTATTGGATACGGAAACCTGAAAATAATGTCTTTCTTTTGATTTT  157296

Qy    1105 ttcccaatcatcaaaaacgtaaaaactactctaggtcgcaaggttaagccattcca      1164
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157295 TTCCCAATCATCAAAAACGTAAAAACTAAAAACTACTCTTAGGTCGTCGCAAGGTTAAGCCATTCTCA  157236

Qy    1165 gcttagcagtggcaggctggcttgtgacctacagttggccaatcccgtgttccc      1224
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157235 GCTTAGCAGTGGCAGGCTGGCTTGTGACCTACAGTTGGCCAATCCCTGATTCCC  157176

Qy    1225 aaatgtattcctcaggagtgtgggcaaatacttatggaagtgctggattaaacagagt      1284
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157175 AAAATGTATTCCTCAGGATGTGTGGGCAAATACTTATGGGAAATGCTGATTAAACAGAGT  157116

Qy    1285 taagaagcatcagacatttccaggacgggctagcacatgccagggctctctaactgacct      1344
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157115 TAAGAAGCATCAGACATTTTCCAGGACGGGCTAGCACATGCCAGGGCTCTCTAACTGACCT  157056

Qy    1345 catt-ggattcatctgtttcatggaggatcttgcaagacaagaattcctcaaacctagag      1403
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157055 CATTAGGATTCATCTGTTTCATGGAGGATCTTGCAAGACAAGAATTCCTCAAACCTAGAG  156996

Qy    1404 tctgaggactgtgctcttggaaacactgctctctgctctgatgccctcactgggcacatggta      1463
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156995 TCTGAGGACTGTGCTCTTGGGAAACACTGCTCTCTGCTCTGATGCCCTCACTGGGCACATGGTA  156936

Qy    1464 gaatctgagagctgagtgcctgctagctgagatagggtcagagtcttgactgcctgg      1523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156935 GAATCTAGAGCTGAGTGCCTGCTAGCTGGAGATAGGGTCAGAGTCTTGACTGCCCTGG  156876

Qy    1524 cagtcttgacatcacagtctgtctgtcgtgtcgtccccgagtggttcagagcacacaggccaag      1583
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156875 CAGTCTTGACACATCACGCTGTCTGTTGTCTGTGTCCCCGAGTGGTTCAGAGCACCACACAGGCCAAG  156816

Qy    1584 actagccacacagagcaccagagcctcaccagtctaaatcacatggttcatttcc      1643
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156815 ACTAGCCACACCAGAGCACCACCAGGCCTCAGCTTTCCCATGCGGCTTTGTCCATGGGCTTACATTTCC  156756

Qy    1644 ttattcttcctggttctccagaacctaaggagaggcacattttggttgagtgattataacc      1703
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156755 TTATTCTTCCTGGTTTCCAGAACCTAAGGAGAGGCACATTTTGGTTGAGTGATTATAACC  156696

Qy    1704 ctaggaccatggtggtagctgcagtgcatgtcggaggaaactcctcaactcctggccctgatgga      1763
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156695 CTAGGACCATGGTAGCTGCAGTGCGCATGTCGCATGCAGAAAACACTCCTCAACTTCCTGGCCCTGATGGA  156636

Qy    1764 ttaaagagaggtacttacaggttatttctgcttgctggactcactgtcccagcacatgaata      1823
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 156635 TTAAAGGAGGAGGTACTTACACAGGTTATTCTTCGCTGTGGACTACTGTCCCAGCATGAATA 156576
           ||||||||||||| |||||||| |||||||||||||| |||||||||||||||||||||||
Qy   1824 gggcatcattattgaattatttgacacaggaaggagagactggtgtatgctgcacagtaataa 1883
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db 156575 GGGCATCATTATTGAATTATTTGACACAGGAAGGAGAGACTGGTGTATGCTGCACAGTAATAA 156516

Qy   1884 tgtatttacatgtgtacagagtttaccaagcacctctgtgttgtttttgcctttgttttat 1943
           ||||||||||||||||| ||| ||||||||||||||||||||||||||||||| ||||| ||
Db 156515 TGTATTTACATGTGTACACAGTTTACCAAGCACCTCTGTGTTGTTTTTGCCTCTGTTTAT 156456

Qy   1944 tacactt--------gggacaaatttttaaaattatacatgcagagact 1985
           ||||||          |  | ||||||  |||| |||||||| |||||||
Db 156455 TACACTTGCCCTTGGGGGGAGGTGGACAAATTTTTAAAAATTTATACATGCCAGAGACT 156396

Qy   1986 gcagcgcagaagctaagagacttgccctgcccacacagccagtggtagagcctgaac 2045
           ||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db 156395 GCAGCGGCAGAAGCTAAGAGACTTGCCCCTGCCCACACAGCCAGTGGTAGAGCCTGAAC 156336

Qy   2046 tcaaacccaggtctcatctcacctcagggggctgcttcccatcgctgtattgtccttaa 2105
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 156335 TCAAACCCAGGTCTCATCTCACCTCAGGGGCTGCTTCCCATCGCTGTATTGTCCTTAA 156276

Qy   2106 agtgatggtgactaggcaatgagtaattctctaggaaagcatgaccaatttccctttc 2165
           |||||||||||||||||||||||||||||||||||||| ||||| |||||||||||||
Db 156275 AGTGATGGGTGACTAGGCAATGAGTAATTCTCTAGGAAAGCATGACCAATTTCCCTTTC 156216

Qy   2166 tccacctccctctttttctccaccctcccccatcagccccatatatgcccaaatc 2225
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db 156215 TCCACCTCCCTCTTTTTCTCCCACCCTCCCCCATCAGCCCCATATATGCCCAAATC 156156

Qy   2226 tccacaaagcctgcttgcctgccgtcaaacctttacttctgaaatgactccacggctgggac 2285
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156155 TCCACAAAGCCTTGCTTGCCTGCCTGCAAACCTTTACTTCTGAAATGACTTCCACGGCTGGGAC 156096

Qy   2286 g 2286
           |
Db 156095 G 156095
```

RESULT   4
AC068122

```
LOCUS       AC068122              180048 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-237C22, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC068122
VERSION     AC068122.3  GI:7923971
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180048)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
REFERENCE   2  (bases 1 to 180048)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (28-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 18, 2000 this sequence version replaced gi:7717167.
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0237C22
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 158944 bases at least Q40
            Consensus quality: 165826 bases at least Q30
            Consensus quality: 169627 bases at least Q20
            Insert size: 167000; agarose-fp
            Insert size: 177348; sum-of-contigs
            Quality coverage: 3.90 in Q20 bases: agarose-fp
            Quality coverage: 3.76 in Q20 bases: sum-of-contigs
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 28 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *      1400: contig of 1400 bp in length
          *      1500: gap of unknown length
          *      3388: contig of 1888 bp in length
          *      3488: gap of unknown length
          *      4879: contig of 1391 bp in length
          *      4979: gap of unknown length
          *      6965: contig of 1986 bp in length
          *      7065: gap of unknown length
          *      9366: contig of 2301 bp in length
          *      9466: gap of unknown length
          *     11957: contig of 2491 bp in length
          *     12057: gap of unknown length
          *     14208: contig of 2151 bp in length
          *     14308: gap of unknown length
          *     17022: contig of 2714 bp in length
          *     17122: gap of unknown length
          *     20061: contig of 2939 bp in length
          *     20161: gap of unknown length
          *     24025: contig of 3864 bp in length
          *     24125: gap of unknown length
          *     26718: contig of 2593 bp in length
          *     26818: gap of unknown length
          *     28936: contig of 2118 bp in length
          *     29036: gap of unknown length
          *     32791: contig of 3755 bp in length
          *     32891: gap of unknown length
          *     38895: contig of 6004 bp in length
          *     38995: gap of unknown length
          *     44306: contig of 5311 bp in length
          *     44406: gap of unknown length
          *     51589: contig of 7183 bp in length
          *     51689: gap of unknown length
          *     58591: contig of 6902 bp in length
          *     58691: gap of unknown length
          *     63442: contig of 4751 bp in length
          *     63542: gap of unknown length
          *     70554: contig of 7012 bp in length
          *     70654: gap of unknown length
          *     77794: contig of 7140 bp in length
          *     77894: gap of unknown length
          *     85841: contig of 7947 bp in length
          *     85941: gap of unknown length
          *     93227: contig of 7286 bp in length
          *     93327: gap of unknown length
          *    100871: contig of 7544 bp in length
          *    100971: gap of unknown length
          *    112141: contig of 11170 bp in length
          *    112241: gap of unknown length
          *    126358: contig of 14117 bp in length
          *    126458: gap of unknown length
          *    141767: contig of 15309 bp in length
          *    141867: gap of unknown length
          *    155927: contig of 14060 bp in length
          *    156027: gap of unknown length
          *    180048: contig of 24021 bp in length.
```

```
FEATURES          Location/Qualifiers
     source       1..180048
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="1"
                  /clone="RP11-237C22"
BASE COUNT    49152 a    39956 c    40116 g    48114 t    2710 others
ORIGIN

Query Match         94.1%;  Score 2151.4;  DB 2;  Length 180048;
Best Local Similarity   98.3%;  Pred. No. 0;
Matches  2244;  Conservative   0;  Mismatches   16;  Indels   23;  Gaps   6;

Qy      25 gactcgatcacctttgaaccagttgctgctccaaagcttgtactcataactaga 84
           ||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   80705 GGCTCGATCACCTTTTGAACCCAGGTCTGCCTGCCTGCCTCCAAAGCTTGTACTCATAACTAGA 80764

Qy      85 ttctcaactgatgtgttgggccaaggttcctagttctctcctctctgaagta 144
           ||||||||||||| |||||||||||||||||||||||| |||||||||||||
Db   80765 TTCTCAACTGATGTGTTGGGCCAAGGTTCCTCAGGTTCTCTCTTCTTCTGAAGTA 80824

Qy     145 ataatgctatgataagctcatcgagggtgaggccagcacatgttgcctgaactatc 204
           ||||||||||||| ||||||||||||||||||||||||||||||| ||||||||||
Db   80825 ATAATGCTATGATAAGCTCATCCGAGGGCTGAGGCCCAGCACATGTTTGCCTGAACTATC 80884

Qy     205 catgttatatgattccttcctgcagacagagtgagctactcacgatcccaggtgtaccctg 264
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   80885 CATGTTATATGATTCCTTCCTGCAGACAGAGTGAGCTACTCACGATCCCAGGTGTACCCTG 80944

Qy     265 aggcagccaaggtgtatccagagtgagctagctcagttccagcctgcccttaacagct 324
           ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   80945 AGGCCAGCCAAGGTGTATCCATGCCTCCTGTTCCAGCCTCCGTCCCTTTAACAGCT 81004

Qy     325 catccacctgcctgcccccg-cctatctcgacagtagtcagctgtctctaggcacatgg 383
           |||||||||||||||||||| || |||||| |||||||| |||||||||||| |||
Db   81005 CATCCCACCTGCCTGCCTGCCCCCGACCTATCTCAGACCAGTAGTCTAGGATTTCAG-GCC 81063

Qy     384 cctggggctcattttcctctcagttcctgcttagctgtctctcctgcttcccactcac 443
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   81064 GCTCGGGGCTCATTTTCCCTCTCAGTTCCTGCTTAGCTGTCTGTCCTGCCTCCCACTCAC 81123

Qy     444 ctattactccagactctca-cctggtcttcttttctgtctcactgccctcttggtgaca 502
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   81124 CTATTACTCCAGCACTCTCATCTCTGGTCTTCTTTTCTGTCTCACTGCCCTCTTGGTGACAT 81183

Qy     503 ctttatctcatagtagttaggggttcttggtaatgccctaaatccacatggtgggga 562
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   81184 CTTTATCTCATAGTAGTTAGGGGTTCTTGGTAATGCCCTAAATCCACATGGTGTGGGA 81243

Qy     563 agggggagtggggggaagagagtgcgctgtgggggctgtgccactctctgggaggtaagac 622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81244 AGGGGGAGTGGGGGAAGAGAGTGCGCTGTGGGGGCTGTGCCCTACTCTGGGAGGGTRAGAC 81303

Qy     623 tcgggccctccaggacaaaggattcaggtggtggtgcagctatagccaagcagactgctg 682
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   81304 TCGGGCCCTCCAGGAACAAAGGATTCAGGCTGCGTGCTGCCAGCTATAGCCAAGCAGCAGACTGCTG 81363

Qy     683 gccaggattgcaaaggtgttggttgcacagagatgggaagagtctgccaaggtgttct 742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81364 GCCAGGGATTGCAAAGGTGGTGGTTGCACAGAGATTGGGAAGAGTCTGCCAAGGTGTGCT 81423

Qy     743 atgggaggggaggggtgttggttgtgccagaagagaatctgcccagaagggtgttct 802
           ||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||
Db   81424 ATGGGAGGGGAGGGGTGTTGGTTGTGCCAGAGTTGGGAAGAGTTGGGGAAGAGTATTTGGGAAGAGT 81483

Qy     803 actcactcctctttcatctccactgagtctgaggcagttatcctgtcccccac 862
           |||||||||||||| || |||||||||||||||||||||||||||||||||||
Db   81484 ACTCACTCTCCTCTTTCTTCATCTCCACTGAGTCTGAGGCAGTTATCTGTCCCCCAC 81543

Qy     863 gtcacattcctactccgtttcccatgcctggaaccccaggttgggcaaactcttcctgtaa 922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81544 GTCACATTCCTACTCCCCGTTTCCCATGCCTGGAACCCCAGGTTGGGCAAACTCTTCCTGTAA 81603


Qy     923 agaaccagacaggaactatttttaggctccgtgtgccatatggtctcagtcacaactactc 982
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81604 AGAACCAGACAGGAACTATTTTAGGCTCGTGTGCCATATGGTCTCAGTCACAACTACTC 81663

Qy     983 atcctgcctctgtagcacgcaaagcaattagcaacaatagcaacacaatatgtgacc 1042
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81664 ATCCTGCCTCTGCAGCACGCAAAGCAAT TAGCAACAATAGTCAACAAACAATATGTGACC 81723

Qy    1043 ccatgaaaacttattattatggatacggaaacctgaaaataatgtcttcttttgatt 1102
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81724 CCATGAAAACTTTATTATTATGGATACGGAAACCTGAAAATAATGTCTTTCTTTTGATT 81783

Qy    1103 tttcccaatcattaaaaacgtaaaaactactcttagtcgcaaggttaagccattct 1162
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81784 TTTTCCCAATCATTAAAAAACGTTAAAAAACTACTCTTAGTCTCGGCAAGGTTAAGCCATTCT 81843

Qy    1163 cagcttagcagtggcaggctggattggttgtgacctacagttgccaatccctgattc 1222
           |||| ||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   81844 CAGCTTAGCAGTGGCAGGCTGGATTTGGCTTGTGACCTACAGTTGGCCAATCCCTGATTC 81903

Qy    1223 ccaaaatgtattcctcagggatgtggcaaatacttatgggaagtgctggattaaacaga 1282
           |||||||||||||||||||||||||||||||||||||| ||||||||||| |||||||||
Db   81904 CCAAAATGTATTCCTCAGGGATGTGGCAAATACTTATGGGAAATGCTGGATTAAACAGA 81963

Qy    1283 gttaagaagcatcagacatttccaggagacgggctagcacatgccaggggctctctaactgac 1342
           |||| |||||||||||||||||||||| ||||||| |||||||||||||||||||||||||
Db   81964 GTTAAGAAGCATCAGACATTTCCAGGACGGGCTAGCACATGCCAGGGCTCTCTAACTGAC 82023

Qy    1343 ctcatt-ggattcatctgtttcatgggaaacactgctctgcttgatgccctcactgggcacatg 1401
           |||||| |||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db   82024 CTCATTAGGATTCATCTGTTTCATGGGAAACACTGCTCTGCTTGATGCCCTCACTGGGCACATG 82083

Qy    1402 agtctgaggactgtgcttggaaacactgctgcttgctgctctcgtgatgccctcactggcacatgg 1461
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   82084 AGTCTGAGGACTGTGCTTTGGGAAACACTGCTGCTTGCTGCTCTCGTGATGCCCTCACTGGGCACATGG 82143

Qy    1462 tagaatctagagctgagtgcttgctagctgcttgtctgctgagataggtcagagctcttgactgccct 1521
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82144 TAGAATCTAGAGCTGAGTGCTTGCTAGCTGCCTTGTCTGCTGAGATAGGTCAGAGCTCTTGACTGCCCT 82203

Qy    1522 ggcagtcttgacacatcacgctgtctgtgtccctgagtggttcagagccacacaggcca 1581
           ||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
Db   82204 GGCAGTCTTGACACATCACGCTGTCTGTGTCCCTGAGTGGTTCAGAGCCAGTCAGAGCCAGGCCA 82263

Qy    1582 agactagccaccagagcgaccaggcctccccagcttcgtgtcctgctacatt 1641
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   82264 AGACTAGCCACCAGAGCGACCAGGCCTCCCCAGCTTTCGTGTCCTGTCCATGCGTACATTT 82323

Qy    1642 cctatctcctcggttccagaacctaagggaggggcacatttggttgagtgattataa 1701
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   82324 CCTTATTCTTCCGGTTTCCCAGTTTCCAGAACCTAAGGAGAGGCACATTTGGTTGGTGAGTGATTATAA 82383

Qy    1702 ccctagggaccatgggtagctgcatgtcaggagaacactcctcaacttcctgcgcctgatg 1761
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||| |||||
Db   82384 CCCTAGGGACCATGGGTAGCTGCATGTCAGGAGAACCTAAGGAGAGGCACATTTGGTTGGTGAGTGATG 82443

Qy    1762 gattaaggagaggtacttacaggttatttctcgctgtggactactgtcccagcatgaat 1821
           |||||||||||||||||||||||||||| |||||||||||||||||| ||||||||||
Db   82444 GATTAAGGAGAGGTACTTACAGGTTATTTTTGACAGGAAGAGGACTTGCTCGCTCGCACAGTAAT 82502

Qy    1822 tagggcatcattattgaattatttgacagaggaaggagactgtgtatgctgcacagtaat 1881
           |||||||||||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db   82503 TAGGGCATCATCATTATTGAATTATTTGACAGGAAGAGGACTTGCTCGCTCGTATGCTGCCACAGTAAT 82562

Qy    1882 aatgtattacatgtacagagttaccaagcacctctgttcctgtcccccaccatgaa 1941
           |||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||
Db   82563 AATGTATTTACATGTACAGAGTTTACCAAGCACCACCTTTCCACGGAGTTGGTTGTTTTTGCCTTCCTGTTTT 82622

Qy    1942 attacactt------------gggacaaatttttaaaattttatacatgcagaga 1983
           ||||||||        ||||| |||||||||||||||||||||||||||||
Db   82623 ATTACACTTGGCCCTTGGCCCTTTGGGGCGAGGTTGGGCACAAATTTTTTAAAAATTTATACATGCAGAGA 82682
```

```
Qy  1984  ctgcagcgcaagagaagctaagagacttgcccctgcccacacagcagtggtagaagcctga  2043
          |||||||  |||||||| |||||||||||||||||||||||||||||||||||||||||||
Db 82683  CTGCAGCGCGAGAGAAGCTAAGAGACTTGCCCCTGCCCACACAGCCAGTGGTAGAGCCTGA 82742

Qy  2044  actcaaaccagtctcatctcacctcaggggctgcttccccatcgctgtattgtcctt  2103
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 82743  ACTCAAACCCAGTCTCATCTCACCTCAGGGGCTGCTTTCCCCATCGCTGTATTGTCCTT 82802

Qy  2104  aaagtgatggtgactaggcaatgagtaattctctaggaaagcaatgaccaaattcccctt  2163
          |||||||||| |||||||||| |||||||||||||||||||||||| ||||||| |||||
Db 82803  AAAGTGATGGGTGACTAGGCATGCATGAGTAATTCTCTAGGAAAGCATGACCAATTTCCCTT 82862

Qy  2164  tctccacctcctcttttctcccaccctccccatcagcccccatatatatgcccaaa  2223
          |||||||||||||||||| ||| |||||||||||||||||||||||| ||||||||
Db 82863  TCTCCACCTCCTCTTTTTCCTCCACCCTCCCCATCAGCCCCCATATATATGCCCAAA 82922

Qy  2224  tctccacaaagccttgctggctgcaaacttactttctgaaatgacttccacggctggg  2283
          |||||||||||||||| || |||||||| ||||||||||||||||||||||||||||
Db 82923  TCTCCACAAAGCCTTGCTTGCCTGCAAACCTTTACTTCTGAAATGACTTCCACGGCTGGG 82982

Qy  2284  acg  2286
          |||
Db 82983  ACG 82985
```

```
     variation       1889
                     /frequency="0.06"
                     /replace="g"
     mRNA            join(1897..2068,2184..2329,3775..3970,4751..4813,
                     5950..6108,6454..6528,7432..8320)
                     /gene="IL24"
                     /product="interleukin 24"
     gene            1897..8320
                     /gene="IL24"
     variation       2049
                     /gene="IL24"
                     /frequency="0.06"
                     /replace="t"
     CDS             join(2286..2329,3775..3970,4751..4813,5950..6108,
                     6454..6528,7432..7515)
                     /gene="IL24"
                     /codon_start=1
                     /product="interleukin 24"
                     /protein_id="AAL34146.1"
                     /db_xref="GI:17104817"
                     /translation="MNFQQRLQSLWTLARPFCPPLLATASQMQMVLPCLGFTLLWS
                     QVSGAQGQEFHFGPCQVKGVVPQKLWEAFWAVKDTMQAQDNITSARLLQQEVLQNVSD
                     AESCYLWHTLLEFYLKTVFKNYHNRTVEVRTLKSFSTLANNFVLIVSOLQPSQENEMF
                     SIRDSAHRRFLLFRRAFKQLDVEAALTKALGEVDILLTWMQKFYKL"
     repeat_region   2324..2640
                     /rpt_family="SINE"
                     /rpt_type=dispersed
     variation       2461
                     /gene="IL24"
                     /frequency="0.90"
                     /replace=""
     variation       2546
                     /gene="IL24"
                     /frequency="0.01"
                     /replace="a"
     misc_feature    2625..3110
                     /gene="IL24"
                     /note="Region not scanned for variation"
     variation       3221
                     /gene="IL24"
                     /frequency="0.01"
                     /replace="c"
     variation       3260
                     /gene="IL24"
                     /frequency="0.37"
                     /replace="t"
     variation       3292
                     /gene="IL24"
                     /frequency="0.01"
                     /replace="a"
     variation       3352
                     /gene="IL24"
                     /frequency="0.01"
                     /replace="t"
     variation       3466
                     /gene="IL24"
                     /frequency="0.01"
                     /replace="a"
     variation       3517
                     /gene="IL24"
                     /frequency="0.02"
                     /replace="a"
     variation       3530
                     /gene="IL24"
                     /frequency="0.05"
                     /replace="a"
     variation       3551
                     /gene="IL24"
                     /frequency="0.99"
                     /replace=""
     variation       3625
                     /gene="IL24"
                     /frequency="0.37"
```

```
RESULT   5
AY062931
LOCUS       AY062931               11114 bp    DNA     linear   PRI 26-NOV-2001
DEFINITION  Homo sapiens interleukin 24 (IL24) gene, complete cds.
ACCESSION   AY062931
VERSION     AY062931.1  GI:17104816
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 11114)
AUTHORS     Rieder,M.J., Armel,T.Z., Carrington,D.P., Chung,M.-W., Lee,K.L.,
            Poel,C.L., Toth,E.J., Yi,Q. and Nickerson,D.A.
TITLE       Direct Submission
JOURNAL     Submitted (13-NOV-2001) Molecular Biotechnology, University of
            Washington, 1705 NE Pacific, Seattle, WA 98195, USA
COMMENT     To cite this work please use: SeattleSNPs. NHLBI Program for
            Genomic Applications, UW-FHCRC, Seattle, WA (URL:
            http://pga.mbt.washington.edu).
FEATURES             Location/Qualifiers
     source          1..11114
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     repeat_region   543..856
                     /rpt_family="MER1_type"
                     /rpt_type=dispersed
     variation       627
                     /frequency="0.01"
                     /replace="a"
     variation       634
                     /frequency="0.10"
                     /replace="c"
     variation       806
                     /frequency="0.05"
                     /replace="g"
     repeat_region   884..925
                     /rpt_family="MER1_type"
                     /rpt_type=dispersed
     variation       1619..1620
                     /frequency="0.07"
                     /replace=""
     repeat_region   1625..1759
                     /rpt_family="SINE"
                     /rpt_type=dispersed
     variation       1627
                     /frequency="0.02"
                     /replace="c"
```

```
variation        /replace="c"
                 4061
                 /gene="IL24"
                 /frequency="0.01"
                 /replace="a"
variation        4069
                 /gene="IL24"
                 /frequency="0.36"
                 /replace="t"
variation        4082
                 /gene="IL24"
                 /frequency="0.02"
                 /replace="c"
repeat_region    4095..4167
                 /rpt_family="SINE"
                 /rpt_type=dispersed
                 4243
variation        /gene="IL24"
                 /frequency="0.37"
                 /replace="a"
                 4254
variation        /gene="IL24"
                 /frequency="0.01"
                 /replace="g"
                 4332
variation        /gene="IL24"
                 /frequency="0.64"
                 /replace=""
                 4585
variation        /gene="IL24"
                 /frequency="0.17"
                 /replace="a"
                 4602
variation        /gene="IL24"
                 /frequency="0.38"
                 /replace="t"
                 4943
variation        /gene="IL24"
                 /frequency="0.02"
                 /replace="a"
                 5020
variation        /gene="IL24"
                 /frequency="0.01"
                 /replace="a"
                 5389
variation        /gene="IL24"
                 /frequency="0.01"
                 /replace=""
                 5392
variation        /gene="IL24"
                 /frequency="0.01"
                 /replace="t"
                 5519
variation        /gene="IL24"
                 /frequency="0.01"
                 /replace="c"
                 5637..5821
repeat_region    /rpt_family="SINE"
                 /rpt_type=dispersed
                 6016
variation        /gene="IL24"
                 /frequency="0.35"
                 /replace="c"
                 6020
variation        /gene="IL24"
                 /frequency="0.02"
                 /replace="g"
                 6037
variation        /gene="IL24"
                 /frequency="0.01"
                 /replace="c"
                 6115..6280
repeat_region    /rpt_family="SINE"


variation        /rpt_type=dispersed
                 6240
                 /gene="IL24"
                 /frequency="0.01"
                 /replace=""
variation        6282
                 /gene="IL24"
                 /frequency="0.37"


Query Match      81.9%;   Score 1872.2;   DB 9;   Length 11114;
Best Local Similarity 98.6%;   Pred. No. 0;
Matches 1916;   Conservative 0;   Mismatches 8;   Indels 19;   Gaps 2;

Qy   363 gtagtctaggattcagctgcctggggtcatttccctcctcagctcctgcttagc   422
             ||||||||||| ||||||| |||||| |||||||||| ||||| |||||||||||
Db     1 GTAGTCTAGGATTTCAGCTGCCCTGGGGCTCATTTCCCTCTCAGCTTCCTGCTTAGC    60

Qy   423 tgtctcctgcccaccactattactccagcactcaccggtcctccttcctgtc     482
             |||||||||  ||  |||  ||||||||||||| ||||| ||||||      ||||
Db    61 TGTCTCCTGCCTCCCACTCACCTATTACTCCAGCACTCTCACCTGTCTCTTTTTTCTGTC   120

Qy   483 tcatcactgcctcttgacatcttatcccatagtagttagggttcctggtaatgc    542
             |||||||||  |||||||||||| ||| ||||||||| |||||||| |||||||||
Db   121 TCATCACTGCCCTCTTGACATCTTTATCTCATAGTAGTTAGTTAGGGTTCTTGGTAATGC   180

Qy   543 cctaaatccacatggtgggaagggggagtggggaagaggagtgcgctgtggggctgtgc    602
             ||||||||||||||||||||| ||||||||||||||| || |||||||| ||| |||||
Db   181 CCTAAATCCACATGGTGGGAAGGGGGAGTGGGGAAGAGAGTGCGCTGGTGGGCTGTGC    240

Qy   603 ctacttctggagggtaagactcgggccctccaggaacaaaggattcaggctggttggcagc   662
             |||||||||||||||||||||||||||||||||||||||  |||||||||||| ||||||
Db   241 CTACTTCTGGAGGGTAAGACTCGGGCCCTCCAGGAACAACAAAGGATTCAGGCTGGTGGCAGC   300

Qy   663 tatagccaagcagactgctgcccaggggattgcaaaggagtatttgttgcttaagaaaa    722
             ||||||||||||||||||| ||||| ||||| |||||||||||| |||| |||||||||
Db   301 TATAGCCAAGCAGACTGCTGGGCCAGGGATTTGCAAAGGAGTATTTTGTTTGCTTAAGAAAA   360

Qy   723 taaacaactgagtatgagagagagggggtgttggtgccagagagattgggaaga     782
             ||||||||  |||||||| | |||  |||||||||||||||||||||||||||
Db   361 TAAACAACACTGAGTATGAGATGGAGGGAGGGGGTGTTGGTGCCAGAGAGATTGGGAAGA   420

Qy   783 gtctgccaagggtgtgttctactcactctcctcttttcttcattctccactgagctggag   842
             ||||||| ||||| |||||||||||||||||| |||||| |||| |||||||||||||||
Db   421 GTCTGCCCAAGGGGTGTGTTCTACTCACTCTCCTCCTCTTTTCTTTCATCTCCACTGAGCTGGAG   480

Qy   843 gcagttatcctgtcccccagtcacattccactctcctcttcccatgcctggaccccaggt   902
             ||||||||||||||||||||||||||||| ||| ||  ||||||||||||||||||||||
Db   481 GCAGTTATCCTGTCCCCCAGTCACATTCCTACTCCCGTTCCCATGCCTGGACCCCAGGT    540

Qy   903 tgggcaaactcttcctgtaaagaaccagacaggaactatttaggctctgtgtgccatat    962
             ||||||||||||||||||||||||||| |||||||||||| |||||||||| |||||||
Db   541 TGGGCAAACTCTTCCTGTAAAGAACCAGAACAGGAACTATTTTAGGCTCTGTGTGTGCCATAT   600

Qy   963 ggtctcagtcacaactactcatctctgccctgctctgtagcacgaaagcaattagcaacaatat  1022
             |||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||||||
Db   601 GGTCTCAGTCACAACTACTCATCTCTGCCCTCTGCTAGCACGAAAGCAATTAGCAACAATAT   660

Qy  1023 gtcaacaaacatatgtgaccccatgaaaacttatttattatggatacggaaacctgaaa   1082
             ||||||||||||||||||||||||||||||| ||||| || |||| ||||||||||||
Db   661 GTCAACAAACATATGTGACCCCATGAAAAACTTTATTTATTTATTTATGATACGGAAACCTGAAA   720

Qy  1083 ataatgtcttctttgattttttcccaatcattaaaaaacgtaaaaactactccttagg    1142
             ||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||||
Db   721 ATAATGTCTTTCTTTCTTTTGATTTTTTCCCCAATCATTAAAAAACGTAAAAACTACTCCTTAGG   780

Qy  1143 tcgcaaggttaagccattctcagcttagcagtcagtggcaggctgatttggcttgtgacctac  1202
             ||||||||||||||||||||||||||||||| ||||||||||||  |||||||||||||||||
Db   781 TCGCAAGGTTAAGCCATTCTCAGCTTAGCAGTGGCAGGCTGGATTTGGCTTGTGACCTAC    840

Qy  1203 agttggccaatccctgattcccaaaatgtattcctcaggatggtgggcaaatacttatgg   1262
             |||||| ||||||||||||||||||||||||||||||| ||| |||||||||||||||||
Db   841 AGTTGGGCCAATCCCTGATTCCCAAAATGTATTCCTCAGGGATGTGGGCAAATACTTATGG    900
```

RESULT 6
AF276916
LOCUS   AF276916   7025 bp   DNA   linear   PRI 20-DEC-2000
DEFINITION Homo sapiens melanoma differentiation association protein 7 gene,
            complete cds.
ACCESSION   AF276916
VERSION   AF276916.1 GI:11907920
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 7025)
AUTHORS   Peat,J., Kube,D., Eskdale,J., Jueliger,S. and Gallagher,G.
TITLE   The human MDA-7 gene
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 7025)
AUTHORS   Peat,J., Kube,D., Eskdale,J., Jueliger,S. and Gallagher,G.
TITLE   Direct Submission
JOURNAL   Submitted (09-JUN-2000) Department of Surgery, University of
            Glasgow, Queen Elizabeth Building, Glasgow Royal Infirmary, Glasgow
            G31 2ER, Scotland
FEATURES   Location/Qualifiers
     source   1..7025
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     mRNA   join(<1804..1847,3285..3480,4261..4323,5460..5618,
                     5964..6038,6942..>7025)
                     /product="melanoma differentiation association protein 7"
     CDS   join(1804..1847,3285..3480,4261..4323,5460..5618,
                     5964..6038,6942..7025)
                     /codon_start=1
                     /product="melanoma differentiation association protein 7"
                     /protein_id="AAG41401.1"
                     /db_xref="GI:11907921"
                     /translation="MNFQQRLQSLWTLARPFCPPLLATASQMQMVLPCLGFTLLLWS
                     QVSGAQQOEFHFGPCQVKGVVPQKLWEAFWAVKDTMQAQDNITSARLLQQEVLQNVSD
                     AESCYLVHTLLEFYLKTVFKNYHNRTVEVRTLKSFSTLANNFVLIVSQLQPSQENEMF
                     SIRDSAHRRFLFRRAFKQLDVEAALTKALGEVDILLTWMQKFYKL"
BASE COUNT   1759 a   1668 c   1759 g   1838 t   1 others
ORIGIN

Query Match   59.7%; Score 1365.2; DB 9; Length 7025;
Best Local Similarity   97.5%; Pred. No. 0;
Matches 1425; Conservative   0; Mismatches   14; Indels   22; Gaps   3;

Qy   848 tatcctgtccccacgtcacattcctactcccgttccatgcctgacccaggtgggc 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TATCCTGTCCCCACGTCACATTCCTACTCCCGTTCCCATGCCTGACCCAGGTTGGGC   60

Qy   908 aaactcttcctgtaaagaaccagacaggaaactatttaggcctcgtgtgccatatggtct 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAACTCTTCCTGTAAAGAACCAGACAGGAAACTATTTTAGGCGCTGTGTGCCATATGGTCT 120

Qy   968 cagtcaacaactactcatctgcctctgtagcagcagaagcaattagcaacaatatgtcaa 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGTCACAACTACTCATCTGCCTCCTCTAGCAGCACGAAAGCAATTAGCAACAATATGTCAA 180

Qy 1028 caaacatgtgaccccatgaaaacttattattatggatacggaaacctgaaaataat 1087
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAAACATATGTGACCCCATGAAAACTTTATTTATTTATTATGGATACGGAAACCTGAAAATAAT 240

Qy 1088 gtctttctttgattttttcccaatcattaaaaacgtaaaaaactactcttaggtcgca 1147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCTTTCTTTTGATTTTTTCCCCAATCATTAAAAAACGTAAAAAACTACTCTTAGGTCGCCA 300

Qy 1148 aggttaagccattctcagcttagcagtggcagcgtggatttggcttgtgacctacagttg 1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGGTTAAGCCATTCTCAGCTTAGCAGTGGCAGCGTGGATTTGGCTTGTGTGACCTACAGTTG 360

Qy 1208 gccaatcccctgattcccaaaatgtattcctcaggaatgtgggcaaatacttatgggaagt 1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1208

Qy 1263 gaagtgctggattaaacagagttaagaagcatcagacatttccaggacgacgggctagcacat 1322
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GAAATGCTGGATTAAACAGAGTTAAGAAGCATCAGACATTTCCAGGACGACCGGCTAGCACAT   960

Qy 1323 gccaggggctcctaactgacctcatt-ggattcatctgttcatggaggatcttgcgaga 1381
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   961 GCCAGGGCCTCTCTAACTGACCTCATTAGGATTCATCTGTTTCATGTGGAGGATCTTGCAAGA 1020

Qy 1382 caagaattcctcaaacctagagtctgaggactgtgctttgggaaacactgctctgcttga 1441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 CAAGAATTCCTCAAAACCTAGAGTCTCGAGGACTGTGCTTTGGGAAACACTGCTCTGCTTGA 1080

Qy 1442 tgccctcactgggcacatggtagaatctgagctgagtgcctgctagctggagtaggg 1501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 TGCCCTCACTGGGCACATGGTAGAATCTAGAGCTGAGTCCTTGCTAGCTGGAGTAGGG 1140

Qy 1502 tcagagctcttgactgccctggcagtcttgacacatcacgctgtctgtccctgagtg 1561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 TCAGAGCTCTTGACTGCCCCTGCAGTCTTGACACATCACGCTGTCTGTCCCCTGAGTG 1200

Qy 1562 gttcagagcacacaggccaagactagccaccagcaccaggcctccagcttctgg 1621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 GTTCAGAGCCACACAGGCCACAGACTAGCCCACCAGCACCAGGCCTCCAGCTTTCTGG 1260

Qy 1622 gcttgtcatgctacattccttattcttcctgtttccagaacctaaggagaggagca 1681
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GCTTGTCCATGCCTACATTCCTTATTCTTCCTGTTTCCAGAACCTAAGGAGAGGCACA 1320

Qy 1682 ttttggtgatgattataacctagggaccatggtagctgcatgtcaggaaacactcc 1741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 TTTTGGTTGAGTGATTATATAACCCTAGGGACCACATGGTCGCATGTGCAGGAAACACTCC 1380

Qy 1742 tcaacttcctgccctgatggattaaagagaggagtacttacagttatttcttcgctg 1801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 TCAACTTCCTGCCCTGATGGATTAAAGAGAGGAGTACTTACAGGTTATTTCTTCGCTGTG 1440

Qy 1802 gactactgtcccagcatgatagggcatcattattgaattattttgacaggaaggagact 1861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 GACTACTGTCCCAGCATGAATAGGGCATCATTATTATGAATTATTTTGACAGGAAGGAGACT 1500

Qy 1862 ggtgtatgctgcacagtaataatgtatttacatgtgtacagagttaccaagcacctctg 1921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 GGTGTATGCTGCCACAGTAATAATGTATTTACATGTGTACAGAGTTTACCAAGCACCTCTG 1560

Qy 1922 tgttgtttttgccttgtttattacactt----------gggacaaatttt 1963
          ||||||||||||||||||||||||||||| | | ||||
Db 1561 TGTTGTTTTTGCCTCGTTTTATTACACTTGCCCTTGCCCTTGGGGGGAGGTGGGACAAATTTTTTA 1620

Qy 1964 aaaatttatacatgcagagactgcagcgcagagactgtagagactgccctgcccaca 2023
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 AAAATTTATACATGCAGAGACTGCAGCGCGAGACTGTAGAGACTGCCCTGCCCCACA 1680

Qy 2024 cagccagtggtagagccctgacctcaaacccagtctcatctcacctcaggggctgcttc 2083
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 CAGCCAGTGGTAGAGCCTGAACCTCAAACCCAGTCTCATCTCACCTCAGGGGCTGCTTTC 1740

Qy 2084 cccatgctgattgtcctaaagtgatggttgactagccaatgaagtaattcctagga 2143
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 CCCATCGCTGTATTGTCCTTAAAGTGATGGTTGACTAGCCAATGAAGTAATTCCTAGGA 1800

Qy 2144 aagcagaccaattccctttcctttcctcaccctccttttcctccacccctccccatcag 2203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1801 AAGCATGACCAATTTCCCTTTCCTCACCTCCTTTTCCTCCACCCCTCCCCCATCAG 1860

Qy 2204 ccccatatatatgcccaaatctcccacaaagccttgcttgcctgcaaacctttacttctg 2263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1861 CCCCCATATATATGCCCAAATCTCCCACAAAGCCTTGCTTGCCTGCAAACCTTTACTTCTG 1920

Qy 2264 aaatgacttcacggctgggacg 2286
          ||||||||||||||||||||||
Db 1921 AAATGACTTCCACGGCTGGGACG 1943

```
RESULT   7
HSA243947
LOCUS       HSA243947           29034 bp    DNA     linear   PRI 09-NOV-2000
DEFINITION  Homo sapiens VCX-A gene for variably charged X chromosome protein.
ACCESSION   AJ243947
VERSION     AJ243947.1  GI:5912555
KEYWORDS    VCX-A gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 29034)
  AUTHORS   Li,X.M., Yen,P.H. and Shapiro,L.J.
  TITLE     Characterization of a low copy repetitive element S232 involved in
            the generation of frequent deletions of the distal short arm of the
            human X chromosome
  JOURNAL   Nucleic Acids Res. 20 (5), 1117-1122 (1992)
  MEDLINE   92195814
REFERENCE   2  (bases 1 to 29034)
  AUTHORS   Fukami,M., Kirsch,S., Schiller,S., Richter,A., Benes,V., Franco,B.,
            Muroya,K., Rao,E., Merker,S., Niesler,B., Ballabio,A., Ansorge,W.,
            Ogata,T. and Rappold,G.A.
  TITLE     A member of a gene family on Xp22.3, VCX-A, is deleted in patients
            with X-linked nonspecific mental retardation
  JOURNAL   Am. J. Hum. Genet. 67 (3), 563-573 (2000)
  MEDLINE   20395828
REFERENCE   3  (bases 1 to 29034)
  AUTHORS   Richter,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-1999) Richter A., Biochemical Instrumentation,
            EMBL, Meyerhofstrasse 1, Heidelberg, 69117, GERMANY
FEATURES             Location/Qualifiers
     source          1..29034
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Xp22.3"
                     /map="between 40BT and DXS1139"
                     /clone="LLNc110"
                     /sex="male"
                     /cell_line="lymphoblastoid"
                     /clone_lib="Lawrence Livermore X-chromosome specific
                     library"
                     /note="Variably charged, X chromosome mRNA on CRI-S232A"
     exon            10574..10732
                     /gene="VCX-A"
                     /number=1
     gene            10574..11947
                     /gene="VCX-A"
     intron          10733..11046
                     /gene="VCX-A"
                     /number=1
     exon            11047..11294
                     /gene="VCX-A"
                     /number=2
     CDS             join(11195..11294,11487..11947)
                     /gene="VCX-A"
                     /codon_start=1
                     /product="VCX-A protein"
                     /protein_id="CAB56197.1"
                     /db_xref="GI:5912556"
                     /translation="MSPKPRASGPPAKATEAGKRKSSQPSPSDPKKKTTKVAKKGKA
                     VRRGRRGKGAATKMAAVTAPEAESGPAAPGPSDQPSQELPQHELPPEPVSEGTOHD
                     PLSQESELEPLSQESEVEEPLSQESQVEEPLSQESEVEEPLSQESQVEEPLSQESEV
                     EEPLSQESQSQVEEPLSQESEMBEELPSV"
     intron          11295..11486
                     /gene="VCX-A"
                     /number=2
     exon            11487..11947
                     /gene="VCX-A"
                     /number=3
     repeat_region   11697..11936
```

```
Db   361  GCCAATCCCTGATTCCCAAAATGTATTCCTCAGGGATGTGGCAAAATACTTATGGGAAAT  420
Qy  1268  gctggattaaacagagttaagaagcatcagacatttccaggacggctagcacatgccag  1327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  GCTGGATTAAACAGAGTTAAGAAGCATCAGACATTTCCAGGACGGCTAGCACATGCCAG   480

Qy  1328  ggctcctaactgacctcatt-ggattcatctgtttcatggaggatcttgcaagacaaga  1386
          ||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db   481  GGCTCTCTAACTGACCTCATTAGGATTCATTCATTGGAGGATCTTGCAAGACAAGA     540

Qy  1387  attcctcaaacctagagtctgagggactgtgcttgggaaacactgctctgatgcccc  1446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  ATTCCTCAAACCTAGAGTCTGAGGACTGTGCTTTGGGAAACACTGCTCTGCTTGATGCCC  600

Qy  1447  tcactggcacatgtagaatctagagctgagtgctgctagctgagataggagtcaga  1506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  TCACTGGCACATGTAGAATCTAGAGCTGAGTGCCTCGTAGCTGAGATAGGGTCAGA     660

Qy  1507  gctcttgactccctggcagtcttgacacatcacgctgtctgtgtcccctgagtggttca  1566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GCTCTTGACTGCCCTGGCAGCTTGACACATCACGCTGTCTGTGTCCCCTGAGTGGTTCA  720

Qy  1567  gagccacacaggccaagactagcccaccagagcaccaggcctccagctttctgggcttg  1626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GAGCCACACAGGCCAAGACTAGCCCACCAGAGCACCAGGCCTCCCAGCTTTCTGGGCTTG  780

Qy  1627  tccatggtacatttccttattcctcctggtttccagaacctaaggagggaggcacatttg  1686
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  TCCATGGTACATTTCCTTATTCCTTATTCCTTCCTGGTTTCCAGAACCTAAGGAGAGGCCACATTTTG  840

Qy  1687  gttgagtgattataacctaggga ccatgggtagctgcatgtcaggaaacactcctcaac  1746
          ||||||||||||||||||      |||||||||||||||||||||||||||||||||
Db   841  GTTGAGTGATTATAACCTAGGGACCATGGTGGGACCATGGTTAGCTGCATGTCAGGAAACACTCCTCAAC   900

Qy  1747  ttcctggccctgatgattaaagaggagggtacttacaggttattcttcgctgt---gga  1803
          |||||||||||||| ||||||||||||||||||||||||||||||||||
Db   901  TTCCTGGCCCTCGTGATGGATAAAGGAGAGGTACTTACACGGTTATTTCTTCGCTATTACGGA   960

Qy  1804  ctactgtcccagcatgaataggcatcattatgaattatttgacaggaaggagactgg  1863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  CTACTGTCCCAGCCATGAATAGGCCATCATCAATATTTGAATTATTTGACAGGAAGGAGACTGG  1020

Qy  1864  tgtatctgacagtaataatgtattacatgtgtacagagtttaccaagcacctctgtg  1923
          |||||||||||||||||||| |||||||||||| |||||||||||||||||||||
Db  1021  TGTATGCTGCACAGTAATAATGTATTTACATGTCGTACAGAGTTTACACAGCACCCTCTGTG  1080

Qy  1924  ttgttttggcctttgcttattacactt----------gggacaaattttaa  1965
          |||||||||||||||||||||||| ||   | |||| ||
Db  1081  TTGTTTTGGCCTCGTGTTTATTACACTTGCCCCTTGGGGGAGTGGGACAAATTTTTTTAAA  1140

Qy  1966  aatttatacatgcaagagactgcacgcgagagagctaaggagacttgccccctgcccacaca  2025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  AATTTATACATGCAAGAGACTGCACGCGAGAGAAGCTAAGAGACTTGCCCCTGCCCCCACACA  1200

Qy  2026  gccagtggtagagcctgaactgaactcaaaccaggtctcatcctcacctcagggggctgctttccc  2085
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  GCCAGTGGTGAGAGCCTGAGACCTGAACTCAAACCCAGGCCTCATCCTCACCTCAGGGGGCTGCTTTCCC  1260

Qy  2086  catcgctgtattgtccttaaagtgatggtgactaggcaatgaagtaattctctaggaaa  2145
          |||||||| |||  ||||||||||||||||||||||||||||||||||||
Db  1261  CATCCGTCGAATTGCCCTTAAAGTGATGATGGGGTGGACTAGGCAATCAAGTAATTCCTAGGGAAA  1320

Qy  2146  gcatgaccaatttcccttctctccacctcctcttttctctccaccctcccccatcagcc  2205
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  GCATGACCAATTTCCCTTCTCCACCTCCCCTCTTTTCTCCACCCTCCCCCATCCAGCC  1380

Qy  2206  cccatatatgcccaaatcccacaaagcctgctgcctgacgaaacctttacttctgaa  2265
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  CCCATATATATGCCCAAAATCCCACAAGCCTTCGTTGCCTGCCTGCCAAACCTTTACTTCTGAA  1440

Qy  2266  atgacttcccacggctgggacg  2286
          ||||||||||||||||||||
Db  1441  ATGACTTCCCACGGCTCGGGACG  1461
```

```
                    /citation=[1]
                    /rpt_family="RU1"
repeat_region    12097..12297
                    /citation=[1]
                    /rpt_family="RU2"
BASE COUNT     9123 a    5623 c    5586 g    8631 t    71 others
ORIGIN

Query Match        4.0%;    Score 91.8; DB 9;   Length 29034; .
Best Local Similarity  75.5%;    Pred. No. 1.7e-14;
Matches  114;  Conservative    0;  Mismatches    37;  Indels   0;  Gaps    0;

Qy   891  ctggaccaggttggcaaactcctcctgtaaagaaccagacaggaactatttaggctc  950
          ||  |   ||| ||| || ||||||||||||||||||||||||||||||||||||||
Db  4201  CTAAATCAGAGTTCTGCAACCTTTCCTGTAAAGGACCAGCAGGAAGTATTTTAGGCTT  4260

Qy   951  tgtgtgccatatggtctcagtcacaactactcatctctgcctctgtagcacgaaagcaat  1010
          |  |   ||||||||||||| | ||||||||||||||||||||| || |||  |||
Db  4261  TGAAGACAATATGGTCTCTTTTACAACTACTCATCTCTGCCTTTGTAGCAAAAAAACAGC  4320

Qy  1011  tagcaacaatatgtcaacaaacatatgtgac  1041
           ||||||||||| ||  ||| ||||| |
Db  4321  CACAGACAATATGTAAATTAATAAATGTGGC  4351


RESULT    8
AC108684/c
LOCUS       AC108684      174210 bp     DNA     linear    HTG 31-JAN-2002
DEFINITION  Homo sapiens chromosome Xp clone RP11-359020, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC108684
VERSION     AC108684.1  GI:18449745
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174210)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Lieu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,N., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
```

```
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu.C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174210)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ------------------ Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----------------- Project Information
            Center project name: HDOM
            Center clone name: RP11-359020
            ----------------- Summary Statistics
            Sequencing vector: plasmid; M77789
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329First call to
            findPhrapList
            Consensus quality: 170951 bases at least Q40
            Consensus quality: 171318 bases at least Q30
            Consensus quality: 171608 bases at least Q20
            Estimated insert size: 170999; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 9.9x in Q20 bases; sum-of-contigs estimation
          * NOTE: Estimated insert size may differ from sequence length
          *  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved
          *      1  76682:  contig of 76682 bp in length
          *  76683  76782:  gap of unknown length
          *  76783 150676:  contig of 73894 bp in length
          * 150677 150776:  gap of unknown length
          * 150777 164461:  contig of 13685 bp in length
          * 164462 164561:  gap of unknown length
          * 164562 174210:  contig of 9649 bp in length.
FEATURES             Location/Qualifiers
     source          1..174210
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Xp"
                     /clone="RP11-359020"
BASE COUNT     52809 a    34893 c    34663 g    51537 t    308 others
ORIGIN

Query Match        4.0%;    Score 91.8; DB 2;   Length 174210;
Best Local Similarity  75.5%;    Pred. No. 1.8e-14;
Matches  114;  Conservative    0;  Mismatches    37;  Indels   0;  Gaps    0;

Qy    891  ctggaccaggttggcaaactcctcctgtaaagaaccagacaggaactatttaggctc  950
           ||  |   ||| ||| || ||||||||||||||||||||||||||||||||||||||
Db 159418  CTAAATCAGAGTTCTGCAACCTTTCCTGTAAAGGACCAGCAGGAAGTATTTTAGGCTT 159359

Qy    951  tgtgtgccatatggtctcagtcacaactactcatctctgcctctgtagcacgaaagcaat 1010
           |  |   ||||||||||||| | ||||||||||||||||||||| || |||  |||
Db 159358  TGAAGACAATATGGTCTCTTTTACAACTACTCATCTCTGCCTTTGTAGCAAAAAAACAGC 159299
```

```
Qy      1011 tagcaacaatatgtcaacaaacatatgtgac 1041
             | ||||||||| ||  || || ||||| |
Db    159298 CACAGACAATATGTAAATTAATAAATGTGGC 159268


RESULT    9
AC002090/c
LOCUS       AC002090              137769 bp    DNA     linear   PRI 06-JUN-1998
DEFINITION  Genomic sequence from Human 17, complete sequence.
ACCESSION   AC002090
VERSION     AC002090.1  GI:2160130
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137769)
AUTHORS     Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,K.H., and Lander,E.S.
TITLE       Genomic sequence from Human 17
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 137769)
AUTHORS     Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,N., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (13-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 137769)
AUTHORS     Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
            Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
            Marquis,N., McDermott,J., Moloney,N., Morrow,N., Nachman,A.,
            Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
            Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 137769)
AUTHORS     Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
            McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
            Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraigery,K.,
            Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
            Lane,M., MacKenzie,J., Marquis,N., McDermott,J., Molla,M.,
            Morrow,J., Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T.,
            Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Wilmer,F., Zemtseva,I. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (02-SEP-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5  (bases 1 to 137769)
AUTHORS     Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
            McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
            Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraigery,K.,
            Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
            Lane,M., MacKenzie,J., Marquis,N., McDermott,J., Molla,M.,
            Morrow,J., Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T.,
            Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Wilmer,F., Zemtseva,I. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (09-SEP-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   6  (bases 1 to 137769)
AUTHORS     Hawkins,T.L., Birren,B.W., Fasman,K.H., Nussbaum,C., Lander,E.S.,
            McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
            Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraigery,K.,
            Hagos,B., Harris,K., Huang,J., Hui,L., Jacotot,L., Kirby,A.,
            Lane,M., MacKenzie,J., Marquis,N., McDermott,J., Molla,M.,
            Morrow,J., Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T.,
            Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Wilmer,F., Zemtseva,I. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (06-JUN-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 4, 1997 this sequence version replaced gi:2085774.
            The Staden databases, finishing information, and all
            chromatographic files used in the assembly of this clone are
            available from our anonymous ftp site.

            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1..137769
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="431I4"
                     /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
                     1-194)"
     repeat_region   9..113
     repeat_region   114..416
                     /rpt_family="MLTIC"
     repeat_region   418..660
                     /rpt_family="AluSx"
     repeat_region   951..1021
                     /rpt_family="MLTIC"
     repeat_region   complement(1659..1701)
                     /rpt_family="(CATA)n"
     repeat_region   2009..2282
                     /rpt_family="(CA)n"
     repeat_region   complement(2599..2700)
                     /rpt_family="LINE2"
     repeat_region   2709..2884
                     /rpt_family="MIR"
     repeat_region   complement(3877..4032)
                     /rpt_family="LINE2"
     repeat_region   4630..4841
                     /rpt_family="LINE2"
     repeat_region   complement(5998..6079)
                     /rpt_family="MIR"
     repeat_region   complement(7528..7829)
                     /rpt_family="(GA)n"
     repeat_region   complement(8692..8879)
                     /rpt_family="AluSx"
     repeat_region   9321..9667
                     /rpt_family="MIR"
     repeat_region   complement(9688..9982)
                     /rpt_family="LINE2"
     repeat_region   9985..10210
                     /rpt_family="AluSg"
     repeat_region   11052..11216
                     /rpt_family="LINE2"
     repeat_region   complement(12296..12393)
                     /rpt_family="MIR"
     repeat_region   12572..12644
                     /rpt_family="(GGGA)n"
     repeat_region   12908..12995
                     /rpt_family="AT_rich"
     repeat_region   complement(14208..14363)
                     /rpt_family="LINE2"
     repeat_region   complement(14577..14686)
                     /rpt_family="LINE2"
     repeat_region   14837..14904
                     /rpt_family="(CAGC)n"
     repeat_region   complement(14906..14963)
                     /rpt_family="LINE2"
```

```
Qy   1032 catatgtgacccatgaaaactttattattggatacggaaacctgaaaataatgtct 1091
          ||||| |||||||||| |||| |||||||||| |||   ||| | ||  |||
Db  44870 TAGGGTGACTGTTCCAATGAAACATTGTTTCAGATGCTGAAAGTTGAATTTTATATAA 44811

Qy   1092 ttct 1095
          |||
Db  44810 TTTT 44807
```

```
RESULT 10
AL139275
LOCUS       AL139275             193126 bp    DNA       linear    PRI 04-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-552E20 on chromosome
            6p12.3-21.2, complete sequence.
ACCESSION   AL139275
VERSION     AL139275.30  GI:15485119
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193126)
AUTHORS     Smith,M.
TITLE       Direct Submission
JOURNAL     Submitted (04-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 2001 this sequence version replaced gi:14800199.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-552E20 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-552E20 The true
            left end of clone RP11-535K1 is at 161600 in this sequence. The
            true right end of clone RP11-200E19 is at 35574 in this sequence.
FEATURES             Location/Qualifiers
     source          1..193126
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p12.3-21.2"
                     /clone="RP11-552E20"
                     /clone_lib="RPCI-11.2"
     misc_feature    12..455
                     /note="match: GSS: Em:AQ807014"
     misc_feature    4253..4780
                     /note="match: GSS: Em:B71338"
     misc_feature    4256..4678
                     /note="match: GSS: Em:AQ437688"
     misc_feature    complement(8999..9566)
```

```
     repeat_region   /rpt_family="LINE2"
                     complement(17782..17841)
     repeat_region   /rpt_family="AT_rich"
                     complement(19230..19258)
     repeat_region   /rpt_family="AT_rich"
                     complement(19446..19742)
     repeat_region   /rpt_family="AluY"
                     19749..19785
     repeat_region   /rpt_family="AT_rich"
                     complement(20413..20518)
     repeat_region   /rpt_family="MER5B"
                     21555..21649
     repeat_region   /rpt_family="MER5A"
                     complement(21855..21892)
     repeat_region   /rpt_family="LINE2"
                     complement(22871..22908)
     repeat_region   /rpt_family="AT_rich"
                     complement(23129..23168)
     repeat_region   /rpt_family="LINE2"
                     23894..24398
     repeat_region   /rpt_family="LINE2"
                     complement(23296..23654)
     repeat_region   /rpt_family="MLT2CA"
                     complement(24536..24675)
     repeat_region   /rpt_family="MIR"
                     complement(24946..25005)
     repeat_region   /rpt_family="MIR"
                     26012..26140
     repeat_region   /rpt_family="LINE2"
                     26342..26477
     repeat_region   /rpt_family="MIR"
                     26712..26929
     repeat_region   /rpt_family="MER58A"
                     complement(27263..27327)
     repeat_region   /rpt_family="MIR"
                     complement(28097..28886)
     unsure          /note="Sequence in this region could not be obtained from
                     the other strand or with an alternative sequencing
                     chemistry.
                     "
     repeat_region   28421..28587
                     /rpt_family="MIR"
     repeat_region   29156..29359
                     /rpt_family="L1"
     repeat_region   30191..30324
                     /rpt_family="MIR"
     repeat_region   30411..30728
                     /rpt_family="MER7A"
     repeat_region   complement(30998..31061)
                     /rpt_family="LINE2"
     repeat_region   complement(31629..31659)
                     /rpt_family="AT_rich"
     repeat_region   33484..34739
                     /rpt_family="LTR12"
     repeat_region   34740..35075
                     /rpt_family="THE1B"
     repeat_region   complement(35105..35297)
                     /rpt_family="MIR"
     repeat_region   complement(35585..35766)
                     /rpt_family="MIR"
     repeat_region   complement(36289..36589)
```

```
Query Match      4.0%;  Score 91.2;  Pred. No. 2.7e-14;  DB 9;  Length 137769;
Best Local Similarity 68.5%;
Matches 126;  Conservative 0;  Mismatches 58;  Indels 0;  Gaps 0;

Qy    912 tctttcctgtaaagaaccagacagagaactatttagctctgtgtgccatatggtctcagt 971
          | ||||  ||  ||| ||||||| ||  ||||||| ||  ||||||||||||||||||||
Db  44990 TTTTCCAGTGAAGGGCCAGATAGTAAACATGTTAGGCTTCGTGGGCCATATGGTCTCAGT 44931

Qy    972 cacaactactcatcctcgcctctgtagcacgaaagcaattgacgcaacaatatgtcaacaaa 1031
          |||||||||| | |||||| |||||| ||    ||| | ||||| ||||||||||| |||||
Db  44930 TGCAACTACTCAACTCTGCCTTTGTAGTGGGAAAGCAGTCACAGACAATATGCAAACAAA 44871
```

misc_feature   /note="match: STS: Em:G63453"
               9829. .10237
misc_feature   /note="match: GSS: Em:AQ030932"
               complement(10116. .10617)
misc_feature   match: GSS: Em:G56076
               13203. .13321
misc_feature   /note="match: GSS: Em:AQ792274"
               13238. .13428
misc_feature   /note="match: GSS: Em:AQ042370"
               complement(13332. .13469)
misc_feature   /note="match: GSS: Em:AQ504765"
               15417. .15794
misc_feature   /note="match: GSS: Em:AQ138303"
               15426. .15971
misc_feature   /note="match: GSS: Em:AQ384582"
               15435. .16067
misc_feature   /note="match: GSS: Em:AQ198235"
               20235. .20758
misc_feature   /note="match: GSS: Em:AQ480601"
               20275. .20620
misc_feature   /note="match: GSS: Em:AQ097268"
               33115. .33591
misc_feature   /note="match: GSS: Em:AQ613320"
               complement(35079. .35571)
misc_feature   /note="match: GSS: Em:AQ692857"
               complement(38680. .39467)
misc_feature   /note="match: GSS: Em:AQ892878"
               41966. .42582
misc_feature   /note="match: GSS: Em:AQ308209"
               41976. .42381
misc_feature   /note="match: GSS: Em:AQ124355"
               46566. .46984
misc_feature   /note="match: GSS: Em:AQ882156"
               47399. .47848
misc_feature   /note="match: GSS: Em:AQ811844"
               64782. .65002
misc_feature   /note="match: GSS: Em:AQ772609"
               64783. .65002
misc_feature   /note="match: GSS: Em:AQ683445"
               complement(64835. .64983)
misc_feature   /note="match: STS: Em:G08189"
               complement(64864. .64983)
misc_feature   /note="match: GSS: Em:AL328221"
               complement(64864. .64975)
misc_feature   /note="match: GSS: Em:AL324222"
               complement(64888. .64983)
misc_feature   /note="match: GSS: Em:AQ932064"
               64898. .65015
misc_feature   /note="match: STS: Em:G09742"
               complement(64898. .64990)
misc_feature   /note="match: GSS: Em:AZ257084"
               complement(64898. .64989)
misc_feature   /note="match: GSS: Em:AQ674454"
               complement(64900. .64983)
misc_feature   /note="match: GSS: Em:AZ237938"
               64903. .65013
misc_feature   /note="match: GSS: Em:AL191898"
               64903. .64991
misc_feature   /note="match: GSS: Em:AL309316 Em:AL316981 Em:AL346021"
               64903. .64983
misc_feature   /note="match: GSS: Em:AL224697"
               complement(64903. .64991)
misc_feature   /note="match: GSS: Em:AL284370"
               complement(64903. .64983)
misc_feature   /note="match: STS: Em:G10109"
               64905. .65002
misc_feature   /note="match: GSS: Em:AQ284087"
               complement(74424. .74751)
misc_feature   /note="match: GSS: Em:AQ151384"
               complement(77154. .77547)
misc_feature   /note="match: GSS: Em:AQ192713"
               80797. .81303

misc_feature   /note="match: GSS: Em:AQ606690"
               complement(80886. .81417)
misc_feature   /note="match: GSS: Em:AQ602297"
               complement(91202. .91756)
misc_feature   /note="match: GSS: Em:AQ063130"
               complement(102120. .102591)
misc_feature   /note="match: GSS: Em:AQ177699"
               107123. .107721
misc_feature   /note="match: GSS: Em:AQ524997"
               complement(115935. .116552)
misc_feature   /note="match: GSS: Em:AQ081194"
               complement(116067. .116478)
misc_feature   /note="match: GSS: Em:AQ031004"
               complement(116195. .116555)
misc_feature   /note="match: GSS: Em:B32716"
               complement(128470. .128806)
misc_feature   /note="match: GSS: Em:B36077"
               complement(128811. .129176)
misc_feature   /note="match: STS: Em:HS164XF4"
               139541. .139825
misc_feature   /note="match: GSS: Em:AQ278260"
               165778. .165946
misc_feature   /note="match: STS: Em:G26303"
               171874. .172092
misc_feature   /note="Single clone region. Sequence from reads from a
               short insert library derived from a clone PCR.
               Restriction digest data confirm the assembly."
               complement(175807. .176203)
misc_feature   /note="match: GSS: Em:AQ882642"
               176202. .176743
misc_feature   /note="match: GSS: Em:AQ374550"
               complement(177911. .178307)
misc_feature   /note="match: GSS: Em:AQ131194"
               complement(177997. .178301)
misc_feature   /note="match: GSS: Em:AQ148743"
               178325. .178821
misc_feature   /note="match: GSS: Em:AQ420001"
               complement(179120. .179217)
misc_feature   /note="match: GSS: Em:AQ545214"

BASE COUNT   54062 a   43323 c   43111 g   52630 t
ORIGIN

Query Match
Best Local Similarity   4.0%;   Score 90.8; DB 9; Length 193126;
Matches 194; Conservative   0; Mismatches 152; Indels   2; Gaps   1;
                                       55.7%; Pred No. 3.5e-14;

Qy   901 gttgggcaaactcttcctgtaaagaaccagacaggaactatttaggctctgtgtgccat 960
             | |||||| |||||| |||||||||||| || |||||||||||| ||| ||||| ||||
Db   125379 GGTCAGCAAACATTTTCGTAAAGAACCAGATAGTGCATATTTTAGGCTTTGTGGGCCAT 125438

Qy   961 atggtccagtcacaactactcatctcgccctgtagcacgaaagcaatagcaacaat 1020
             || ||| ||| |||||| |||||||||||||||| |||| |||||||| ||| |||
Db   125439 ATAATCTGTGCACCACTGCTCACCTCCGCCATTGCACTGTGAAAGCAGCTGTAGACAAA 125498

Qy   1021 atgtcaacaaacatatgtgaccccatgaaaacttatttattatggatacggaaacctga 1080
             | || ||| | ||| ||| |||||| || ||| || || || |||||| ||||
Db   125499 ACGTAAATGAACGGGCATGGGCTGTGCCCCAGTAAAACTTTATGGGCACCATGAAATTTAA 125558

Qy   1081 aaataatgtctttctttgatttttccccaatcattaaaaacgtaaaaactactcctta 1140
             | | ||| ||||| ||||| |||| |||||| |||||| ||||||| |||||| ||
Db   125559 TTTTCATGTGTCACAAAGTATTTTTCTTTGATTTTCCAACCAATTAAAATGTAAAAAC 125618

Qy   1141 ggtcgcaaggtaagccattccagcttagcagtggcaggctggattggcttgtgacct 1200
             | ||| | || | | || ||||| | ||||||||||||||| ||||| |||
Db   125619 CTTTCTTAACTTGAGGGTAGAACAAAACAG--GCATCGGGGTCGGGATTTGGCTCATGGGCC 125676

Qy   1201 acagttggccaatccctgattcccaaaatgtattcctcagggatgtgg 1248
             | || | ||||||||||||| | ||||||||| ||||| ||||||
Db   125677 ATTGTGTGCCGACCCCCTGATTGACATCATCCCCATCTCACAAATGTGG 125724

RESULT 11

```
AC025298
LOCUS       AC025298      177957 bp    DNA    linear   HTG 03-JUN-2000
DEFINITION  Homo sapiens clone RP11-2N21, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   AC025298
VERSION     AC025298.3  GI:8225254
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177957)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Unpublished
  JOURNAL   Homo sapiens, clone RP11-2N21
REFERENCE   2  (bases 1 to 177957)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkly,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,C., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 3, 2000 this sequence version replaced gi:7670170.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -----------------Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            -----------------Project Information
              Center project name: L2680
              Center clone name: 2_N_21
            -----------------Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 174723 bases at least Q40
              Consensus quality: 176598 bases at least Q30
              Consensus quality: 177197 bases at least Q20
              Insert size: 182000; agarose-fp
              Insert size: 177557; sum-of-contigs
              Quality coverage: 6.7 in Q20 bases; agarose-fp
              Quality coverage: 6.8 in Q20 bases; sum-of-contigs
            -----------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     1828: contig of 1828 bp in length
            *   1829     1928: gap of      100 bp
            *   1929    26417: contig of 24489 bp in length
            *  26418    26517: gap of      100 bp
            *  26518    54437: contig of 27920 bp in length
            *  54438    54537: gap of      100 bp
            * 101425   101524: gap of      100 bp
            * 101525   177957: contig of 76433 bp in length.
FEATURES             Location/Qualifiers
     source          1..177957
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-2N21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1828
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:right"
     misc_feature    1929..26417
                     /note="assembly_fragment"
     misc_feature    26518..54437
                     /note="assembly_fragment"
                     clone_end:SP6
                     vector_side:right"
     misc_feature    54538..101424
                     /note="assembly_fragment"
     misc_feature    101525..177957
                     /note="assembly_fragment"
BASE COUNT    53204 a   34244 c   35114 g   54995 t    400 others
ORIGIN

Query Match        3.9%;  Score 90;  DB 2;  Length 177957;
Best Local Similarity  57.1%;    Pred. No. 5.9e-14;
Matches  186;   Conservative    0;   Mismatches 135;   Indels    5;   Gaps   1;

Qy    829 ccactgagctgaggcagttatcctgtccccacgtcacattcctactccgttcccat 888
          ||  |||||  |||| ||   |||||| || |  |  ||| | | |||||||
Db  52843 CCCTAAAATGGAAATGGAATTGATCCCACCCAGACCCACCTTGGGAGAGATGGTTTCT 52902

Qy    889 gcctggaccaggttgggcaaactcttcctgtaaagaaccagacgaggaacattttaggc 948
          ||||| ||| |||||| ||| ||  | ||||| |||| |  ||||| ||||||||||||
Db  52903 AATGGGACCAGGATCAGCAAACTTTCCGGTGAAAGACCACACAGTAAATATTTTAGGC 52962

Qy    949 tctgtgtgccatatggtctcagtcacaactacctcatctgcctctgtagcacgaaagca 1008
          |  |||| |||| | ||| |||  | |||||||| ||  ||||||| ||  | ||| |||||
Db  52963 TTGTGGGTCATACGGTTTCGTTGCAGCTACCTGCAGCTACCTCGCCATTGCATCATGAAAGCA 53022

Qy   1009 attagcaacaaatgtcaacaaacatatgtg----accccatgaaaactttattatta 1063
          |||||| | |||| ||| ||  ||||||||     |||||||||||||||||||||
Db  53023 ACCAGAGACGCTAAGTAAGTAAATGAATGAGCATGGCTGTGCTCCAAAAAAACTTTACAGACAC 53082

Qy   1064 tggatacggaaactgaaataatgtctttctttgattttttcccatcattaaaaaa 1123
          || ||| | ||  |  |||| ||| ||| |||  ||  |||||||||||||||||
Db  53083 TGAAATTTGAATTTATATAATTTCAGGTATCATGAAATATTCTTTAACCACTAAAATA 53142

Qy   1124 cgtaaaaactactcttagtcgcaag 1149
          || |||||| ||| | ||  |||||
Db  53143 TGTAAAATCATTCTTAGCTCATAGG 53168

RESULT 12
AC022801/c
LOCUS       AC022801      178477 bp    DNA     linear   HTG 17-MAR-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-354A24 map 14, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC022801
VERSION     AC022801.3  GI:7259748
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1 (bases 1 to 178477)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 14, clone RP11-354A24
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 178477)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Chopel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 17, 2000 this sequence version replaced gi:6980319.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L6051
            Center clone name: 354_A_24
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 172070 bases at least Q40
            Consensus quality: 175121 bases at least Q30
            Consensus quality: 176261 bases at least Q20
            Insert size: 181000; agarose-fp
            Insert size: 177277; sum-of-contigs
            Quality coverage: 5.4 in Q20 bases: agarose-fp
            Quality coverage: 5.5 in Q20 bases: sum-of-contigs
            --------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 13 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1   1693: contig of 1693 bp in length
          *   1694  1793: gap of    100 bp
          *   1794  4078: contig of 2285 bp in length
          *   4079  4178: gap of    100 bp
          *   4179 14243: contig of 10065 bp in length
          *  14244 14343: gap of    100 bp
          *  14344 21787: contig of 7444 bp in length
          *  21788 21887: gap of    100 bp
          *  21888 31116: contig of 9229 bp in length
          *  31117 31216: gap of    100 bp
          *  31217 39038: contig of 7822 bp in length
          *  39039 39138: gap of    100 bp
          *  39139 46800: contig of 7671 bp in length
          *  46801 46909: gap of    100 bp
          *  46910 57558: contig of 10649 bp in length
          *  57559 57658: gap of    100 bp
          *  57659  73890: contig of 16232 bp in length
          *  73891  73990: gap of    100 bp
          *  73991  96635: contig of 22645 bp in length
          *  96636  96735: gap of    100 bp
          *  96736 120417: contig of 23682 bp in length
          * 120418 120517: gap of    100 bp
          * 120518 149372: contig of 28855 bp in length
          * 149373 149472: gap of    100 bp
          * 149473 178477: contig of 29005 bp in length.
FEATURES             Location/Qualifiers
     source          1..178477
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-354A24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1693
                     /note="assembly_fragment"
     misc_feature    1794..4078
                     /note="assembly_fragment"
     misc_feature    4179..14243
                     /note="assembly_fragment"
     misc_feature    14344..21787
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    21888..31116
     misc_feature    31217..39038
                     /note="assembly_fragment"
     misc_feature    39139..46809
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    46910..57558
                     /note="assembly_fragment"
     misc_feature    57659..73890
                     /note="assembly_fragment"
     misc_feature    73991..96635
                     /note="assembly_fragment"
     misc_feature    96736..120417
                     /note="assembly_fragment"
     misc_feature    120518..149372
                     /note="assembly_fragment"
     misc_feature    149473..178477
                     /note="assembly_fragment"
BASE COUNT    51444 a  37184 c  36608 g  52041 t   1200 others
ORIGIN

Query Match             3.9%;  Score 90;  DB 2;  Length 178477;
Best Local Similarity   61.5%;  Pred. No. 5.9e-14;
Matches 182;  Conservative    0;  Mismatches 105;  Indels    9;  Gaps    2;

Qy    861 agctcacattcctactccgttcccatgcctgaccaggtggcaaactcttcctgt 920
          | ||| |||| ||| ||| ||||||| ||| |||| ||| ||||  |||||| |||||
Db  10108 AGGTTCTATTGCTAGCTTGCTACAAATGACCAAGAACAAAGGTCAGTAAACTTTTTCTCT 10049

Qy    921 aaagaaccagacaggaactatttaggctctgtgtccatatggctcagtcacaactac 980
          || || |||| || || |||||||| ||  ||| ||||||| |  || |||| ||||
Db  10048 GAAGGACTAGATAGTAAATATTTCAGGCTTTGCAGACCATATGGTGCCTCTTGCAACTAC 9989

Qy    981 tcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatagtga 1040
          ||| ||||| ||||| ||| || |||| |  ||| ||| |  ||||| ||||| ||||
Db   9988 TCAAGTCTGTCGCTGTGTAGTATGATGAAAGCCACCGTGGACAATATGTAAATGAATGGGGCATGA 9929

Qy   1041 cccat-gaaacttattattatggatacggaaacctgaaaataatgtctttc---- 1094
          | ||| ||| | |||  ||| ||| |  ||| || |||||  |  ||||||
Db   9928 CTGCATTCCAATAAAATTTCTTATGCACATTGAGATTTGAATTTTTACATGTCAGAA 9869

Qy   1095 ---ttttgatttttcccaatcattaaaaacgtaaaaactactcctaggtcgca 1147
          || |||||||||||||| ||| ||| | |  ||||  ||||||
```

Db   9868 AATATTATTCGTTTTCCAATCATTAAAACATATAAAAACCACTCTTAAGTCACA 9813

```
RESULT 13
AC098645
LOCUS       AC098645          192695 bp    DNA     linear   HTG 01-FEB-2002
DEFINITION  Papio cynocephalus anubis clone RP41-470J23, WORKING DRAFT
            SEQUENCE, 5 ordered pieces.
ACCESSION   AC098645
VERSION     AC098645.2  GI:18464060
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      olive baboon.
ORGANISM    Papio cynocephalus anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 192695)
AUTHORS     Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
TITLE       NISC Comparative Sequencing Initiative
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 192695)
AUTHORS     Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            On Feb 1, 2002 this sequence version replaced gi:16506407.
COMMENT     ----------------- Genome Center
            Center: NISC Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            ----------------- Project Information
            Center project name: csb
            Center clone name: 470J23
            -------------------

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ----------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 191421 bases at least Q40
            Consensus quality: 191690 bases at least Q30
            Consensus quality: 191951 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 192295; sum-of-contigs
            Quality coverage: 10.92x in Q20 bases; agarose-fp
            Quality coverage: 10.33x in Q20 bases; sum-of-contigs
            -------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 5 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submitor.
```

```
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved
          *          31167: contig of 31167 bp in length
          *    31168  31267: gap of unknown length
          *    31268  72728: contig of 41461 bp in length
          *    72729  72828: gap of unknown length
          *    72829 100592: contig of 27764 bp in length
          *   100593 100692: gap of unknown length
          *   100693 159516: contig of 58824 bp in length
          *   159517 159616: gap of unknown length
          *   159617 192695: contig of 33079 bp in length.
FEATURES             Location/Qualifiers
     source          1..192695
                     /organism="Papio cynocephalus anubis"
                     /db_xref="taxon:9555"
                     /clone="RP41-470J23"
     misc_feature    1..31167
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1..65403
                     /note="clone overlaps with GenBank Accession Number
                     AC097227 clone RP41-327F23 (center project name csa)"
     misc_feature    31268..72728
                     /note="assembly_fragment"
     misc_feature    72829..100592
                     /note="assembly_fragment"
     misc_feature    100693..159516
                     /note="assembly_fragment"
     misc_feature    147682..192695
                     /note="clone overlaps with GenBank Accession Number
                     AC098698 clone RP41-177A23 (center project name csc)"
     misc_feature    159617..192695
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    58595 a  40608 c  37486 g  55595 t    411 others
ORIGIN
```

```
Query Match              3.9%;  Score 90;  DB 2;  Length 192695;
Best Local Similarity 62.0%;  Pred. No. 5.9e-14;
Matches 160;  Conservative  0;  Mismatches  95;  Indels  3;  Gaps  1;

Qy      884 cccatgcctggacccaggttggcaaactcttcctgtaaagaaccagacaggaactattt 943
            ||--||||||||||||  ||||||||  |  || |||||||  ||||||||| |  |||||
Db   138355 CCGCAGCCTAGAGTAAAAGGTGGCAAAGTTCTCTTGTAAAGAGCCAGACAGTGAATATTT 138414

Qy      944 taggctctgtgtgccatatggtctcagtcaccaactactcatctctgcctctgtagcacga 1003
            ||||| |||  ||  ||  |||||||||||  ||||||||||||||||||  |||||||||
Db   138415 TTGGCTTCGTCGGCCCGATCAGCTCTATCACAACTACTCAACTCTGCCATTGTGAGCACAA 138474

Qy     1004 aagcaattagcaacaatagtcaacaaacatatgtgaccccatgaaaacttttattatta 1063
            ||||| ||  || ||  || || |  |||  |  || |||||||| | ||  | ||||||
Db   138475 AAGCAGTCACAGACGGTATGAAAATAATGAGTATGACTGTGTTCCAATAAAACTTTATTT 138534

Qy     1064 tggatacggaaacctgaaaataatgtcttctttgattttcccaatcattaaaaaa 1123
            ||  ||||  |||  |  |||  |  ||||  ||||||  ||||  |||| |||||
Db   138535 ATGGACGTGAAATCTGAGTTGCAAATAATTC---TCACATTTTTCCAACCATTAAAAA 138591

Qy     1124 cgtaaaaactactcttag 1141
            ||||||||||||||||||
Db   138592 TGTAAAAGCTATTCTTAG 138609
```

```
RESULT 14
AC011060
LOCUS       AC011060          181009 bp    DNA     linear   HTG 12-MAR-2000
DEFINITION  Homo sapiens clone RP11-11G16, WORKING DRAFT SEQUENCE, 25 unordered
            pieces.
ACCESSION   AC011060
```

```
VERSION     AC011060.5  GI:7229784
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181009)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-11G16
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 181009)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkly,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (30-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6984405.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------------------  Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----------------------
            Center project name: L3156
            Center clone name: 11_G_16
            ----------------------  Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 163981 bases at least Q40
            Consensus quality: 172073 bases at least Q30
            Consensus quality: 175479 bases at least Q20
            Insert size: 155000; agarose-fp
            Insert size: 178609; sum-of-contigs
            Quality coverage: 5.1 in Q20 bases; agarose-fp
            Quality coverage: 4.5 in Q20 bases; sum-of-contigs
            ----------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 25 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   1504: contig of 1504 bp in length
          *     1505 1604: gap of    100 bp
          *     1605 2791: contig of 1187 bp in length
          *     2792 2891: gap of    100 bp
          *     2892 3909: contig of 1018 bp in length
          *     3910 4009: gap of    100 bp
          *     4010 5556: contig of 1547 bp in length
          *     5557 5655: gap of    100 bp
          *     5656 7482: contig of 1826 bp in length
          *     7483 7582: gap of    100 bp
          *     7583 8926: contig of 1344 bp in length
          *     8927 9026: gap of    100 bp
          *     9027 10470: contig of 1444 bp in length
          *    10471 10570: gap of    100 bp
          *    10571 12199: contig of 1629 bp in length
          *    12200 12299: gap of    100 bp
          *    12300 14087: contig of 1788 bp in length
          *    14088 14187: gap of    100 bp
          *    14188 15707: contig of 1520 bp in length
          *    15708 15807: gap of    100 bp
          *    15808 17945: contig of 2138 bp in length
          *    17946 18045: gap of    100 bp
          *    18046 20017: contig of 1972 bp in length
          *    20018 20117: gap of    100 bp
          *    20118 22115: contig of 1998 bp in length
          *    22116 22215: gap of    100 bp
          *    22216 24556: contig of 2341 bp in length
          *    24557 24656: gap of    100 bp
          *    24657 27568: contig of 2912 bp in length
          *    27569 27668: gap of    100 bp
          *    27669 32527: contig of 4859 bp in length
          *    32528 32627: gap of    100 bp
          *    32628 37892: contig of 5265 bp in length
          *    37893 37992: gap of    100 bp
          *    37993 41265: contig of 3273 bp in length
          *    41266 41365: gap of    100 bp
          *    41366 52400: contig of 11035 bp in length
          *    52401 52500: gap of    100 bp
          *    52501 63506: contig of 11006 bp in length
          *    63507 63606: gap of    100 bp
          *    63607 79362: contig of 15756 bp in length
          *    79363 79462: gap of    100 bp
          *    79463 97023: contig of 17561 bp in length
          *    97024 97123: gap of    100 bp
          *    97124 124426: contig of 27303 bp in length
          *   124427 124526: gap of    100 bp
          *   124527 149337: contig of 24811 bp in length
          *   149338 149437: gap of    100 bp
          *   149438 181009: contig of 31572 bp in length.
FEATURES             Location/Qualifiers
     source          1..181009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="RP11-11G16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1504
                     /note="assembly_fragment"
     misc_feature    1605..2791
                     /note="assembly_fragment"
     misc_feature    2892..3909
                     /note="assembly_fragment"
     misc_feature    4010..5556
                     /note="assembly_fragment"
     misc_feature    5657..7482
                     /note="assembly_fragment"
     misc_feature    7583..8926
                     /note="assembly_fragment"
     misc_feature    9027..10470
                     /note="assembly_fragment"
     misc_feature    10571..12199
                     /note="assembly_fragment"
     misc_feature    12300..14087
                     /note="assembly_fragment"
     misc_feature    14188..15707
                     /note="assembly_fragment"
     misc_feature    15808..17945
                     /note="assembly_fragment"
     misc_feature    18046..20017
                     /note="assembly_fragment"
     misc_feature    20118..22115
                     /note="assembly_fragment"
     misc_feature    22216..24556
                     /note="assembly_fragment"
     misc_feature    24657..27568
                     /note="assembly_fragment"
     misc_feature    27669..32527
                     /note="assembly_fragment"
```

```
misc_feature    /note="assembly_fragment"
                32628. .37892
misc_feature    /note="assembly_fragment"
                37993. .41265
misc_feature    /note="assembly_fragment
                clone_end:SP6
                vector_side:right"
misc_feature    41366. .52400
                /note="assembly_fragment"
misc_feature    52501. .63506
                /note="assembly_fragment"
misc_feature    63607. .79362
                /note="assembly_fragment
                clone_end:T7
                vector_side:right"
misc_feature    .79463. .97023
                /note="assembly_fragment"
misc_feature    97124. .124426
                /note="assembly_fragment"
misc_feature    124527. .149337
                /note="assembly_fragment"
misc_feature    149438. .181009
                /note="assembly_fragment"

BASE COUNT    57610 a  36532 c  33173 g  51292 t    2402 others
ORIGIN

Query Match         3.9%;     Score 89.2;  DB 2;     Length 181009;
Best Local Similarity  64.6%;  Pred. No. 9.9e-14;
Matches 133;  Conservative   0;  Mismatches   73;  Indels   0;  Gaps   0;

Qy    891 ctggacccaggttgggcaaacctcttcctgtaaagaaccagacaggaacattttaggctc 950
          ||| |||||||  |||||| ||||||| ||||||| |||| |||| | |||||||||
Db 151319 CTiAGCACAGAGGTCAGCAAACTTCTCTGTAAAGGACCAGATGATAAATATTTTAGGTTT 151378

Qy    951 tgtgtgccatatggtctcagtcacacaactactcatctctgcctctgtagcacgaaagcaat 1010
          |||| ||||||||||||||| |||||||||||| |||||| |||| | ||||||| |
Db 151379 TGTGGGCCATATGGTCTCTGTCACAACTACTCCACTCCTCTGCTGTTCTTAGGGCAAAAGCCAGT 151438

Qy   1011 tagcaacaatagtcaacaaacatatgtgacccatgaaaactttattattatggatac 1070
          | ||||  || |||| || || ||||| |||||| |  |  ||||||||||
Db 151439 TATAGAAAATACATAAATGAATGAGCATGGCGTGTGTTCCCATAAAAGTATTTATCGACAC 151498

Qy   1071 ggaaacctgaaaataatgtctttctt 1096
          || |||||  |||| |||| || |||
Db 151499 TGAAATTTGAATTTCATGTAATTTTT 151524

RESULT 15
AL161628
LOCUS       AL161628            183483 bp    DNA        linear      PRI 16-JAN-2001
DEFINITION  Human DNA sequence from clone RP11-31K16 on chromosome 9. Contains
            a snoRNA binding domain pseudogene, the ELAVL2 gene for ELAV
            (embryonic lethal, abnormal vision, Drosophila)-like 2, ESTs, STSs,
            GSSs and a CpG island, complete sequence.
ACCESSION   AL161628
VERSION     AL161628.9  GI:10129841
KEYWORDS    HTG; CpG island; ELAVL2; snoRNA binding domain.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183483)
AUTHORS     Ramsay,H.
TITLE       Direct Submission
JOURNAL     Submitted (20-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 14, 2000 this sequence version replaced gi:10039694.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
```

```
                    corresponding to the overlapping clone, as we submit sequences with
                    only a small overlap as described above.
                    The following abbreviations are used to associate primary accession
                    numbers given in the feature table with their source databases:
                    Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; WP:, WORMPEP; Information
                    on the WORMPEP database can be found at
                    http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
                    was generated from part of bacterial clone contigs of human
                    chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
                    Group.  Further information can be found at
                    http://www.sanger.ac.uk/HGP/Chr9
                    This sequence is the entire insert of clone RP11-31K16 The true
                    left end of clone RP11-321L2 is at 139671 in this sequence. The
                    true right end of clone RP11-315I4 is at 73045 in this sequence.
                    This sequence has been finished according to sequence map criteria
                    as follows.  An attempt is made to resolve all sequencing problems,
                    such as compressions and repeats, but not necessarily within known
                    annotated repeat sequence elements.  Where the sequence is
                    ambiguous, there is an annotation using the 'unsure' feature key.
                    RP11-31K16 is from the library RPCI-11.1 constructed by the group
                    of Pieter de Jong. For further details see
                    http://www.chori.org/bacpac/home.htm
                    VECTOR: pBACe3.6.

FEATURES             Location/Qualifiers
     source          1. .183483
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-31K16"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .224
                     /note="match: GSS: Em:AQ040596"
     misc_feature    1. .456
                     /note="match: GSS: Em:AQ080076"
     misc_feature    24. .653
                     /note="match: GSS: Em:AQ045921"
     repeat_region   415. .471
                     /note="AluJo repeat: matches 60. .118 of consensus"
     repeat_region   472. .763
                     /note="AluY repeat: matches 10. .288 of consensus"
     repeat_region   764. .954
                     /note="AluJo repeat: matches 118. .297 of consensus"
     repeat_region   1096. .1173
                     /note="MLT1J repeat: matches 103. .187 of consensus"
     repeat_region   2555. .4081
                     /note="L1MC2 repeat: matches 4721. .6314 of consensus"
     repeat_region   4184. .4223
                     /note="MLT1I repeat: matches 238. .273 of consensus"
     repeat_region   4424. .4479
                     /note="MLT1I repeat: matches 343. .398 of consensus"
     repeat_region   5387. .5466
                     /note="MER58 repeat: matches 2214. .2294 of consensus"
     repeat_region   5639. .5786
                     /note="LIME3A repeat: matches 6017. .6156 of consensus"
     repeat_region   6111. .6324
                     /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region   6330. .6457
                     /note="MER5A repeat: matches 2. .120 of consensus"
     repeat_region   6840. .7315
                     /note="L2 repeat: matches 2252. .2736 of consensus"
     repeat_region   7345. .7434
                     /note="MIR repeat: matches 2. .89 of consensus"
     repeat_region   7437. .7741
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   7808. .7989
                     /note="MER58 repeat: matches 28. .2412 of consensus"
     repeat_region   10366. .10649
                     /note="MER58B repeat: matches 46. .341 of consensus"
     repeat_region   11076. .11221
                     /note="MLT2 repeat: matches 1. .147 of consensus"
     repeat_region   11222. .11661
                     /note="MIT2CB repeat: matches 1. .504 of consensus"
     repeat_region   11664. .11992
```

```
repeat_region      /note="MLT2 repeat: matches 145. .510 of consensus"
                   12210. .12467
misc_feature       /note="L2 repeat: matches 2482. .2750 of consensus"
                   complement(12696. .13167)
misc_feature       /note="match: GSS: Em:AQ549073"
                   complement(13792. .14266)
misc_feature       /note="match: GSS: Em:AQ594257"
                   complement(14438. .14735)
repeat_region      /note="match: GSS: Em:AQ128238"
                   14756. .14785
repeat_region      /note="15 copies 2 mer aa 100% conserved"
                   16212. .16363
repeat_region      /note="FRAM repeat: matches 4. .145 of consensus"
                   16364. .16385
misc_feature       /note="11 copies 2 mer tt 100% conserved"
                   16462. .16951
repeat_region      /note="match: GSS: Em:AQ631149"
                   16908. .17128
misc_feature       /note="MIR repeat: matches 34. .261 of consensus"
                   17116. .17552
misc_feature       /note="match: GSS: Em:AQ133240"
                   17546. .18050
misc_feature       /note="match: GSS: Em:AQ777145"
                   17660. .17960
repeat_region      /note="match: GSS: Em:AQ094704"
                   19777. .19993
misc_feature       /note="MER58A repeat: matches 2. .224 of consensus"
                   20817. .21173
misc_feature       /note="match: GSS: Em:AQ376698"
                   complement(20820. .20939)
misc_feature       /note="match: GSS: Em:AQ229933"
                   complement(20830. .21133)
misc_feature       /note="match: GSS: Em:AQ235822"
                   complement(20830. .21137)
repeat_region      /note="match: GSS: Em:B65379"
                   21100. .21154
repeat_region      /note="11 copies 5 mer aaaac 78% conserved"
                   21183. .21410
repeat_region      /note="MER30 repeat: matches 1. .230 of consensus"
                   21647. .21773
misc_feature       /note="MIR repeat: matches 119. .246 of consensus"
                   complement(23139. .23500)
repeat_region      /note="match: GSS: Em:AQ210496"
                   23363. .23601
misc_feature       /note="MIR repeat: matches 2. .256 of consensus"
                   complement(24372. .24810)
repeat_region      /note="match: GSS: Em:AQ178605"
                   26871. .27202
repeat_region      /note="MSTD repeat: matches 1. .341 of consensus"
                   27203. .27695
repeat_region      /note="L1PA10 repeat: matches 5642. .6165 of consensus"
                   27696. .27750
repeat_region      /note="MSTD repeat: matches 341. .394 of consensus"
                   27867. .27971
repeat_region      /note="MER5A repeat: matches 84. .189 of consensus"
                   27986. .28150
repeat_region      /note="MER5A repeat: matches 3. .183 of consensus"
                   28765. .29329
misc_feature       /note="match: GSS: Em:B93338"
                   28777. .29151
misc_feature       /note="match: GSS: Em:AQ125799"
                   29425. .29482
repeat_region      /note="29 copies 2 mer ac 87% conserved"
                   29740. .29805
repeat_region      /note="33 copies 2 mer ca 71% conserved"
                   29763. .29807
repeat_region      /note="9 copies 5 mer aaaac 82% conserved"
                   30205. .30412
repeat_region      /note="MIR repeat: matches 2. .230 of consensus"
                   30582. .30732
repeat_region      /note="MER5A repeat: matches 8. .187 of consensus"
                   30911. .31378
misc_feature       /note="match: GSS: Em:AQ546895"


repeat_region      32009. .32080
                   /note="MIR repeat: matches 132. .206 of consensus"
gene               32125. .32784
                   /gene="bA31K16.1"
CDS                32125. .32784
                   /gene="bA31K16.1"
                   /note="bA31K16.1 (snoRNA binding domain pseudogene)
                   match: proteins: Sw:O00567 Tr:Q9VM69 Tr:Q92RW0 Tr:Q9VD11
                   Tr:O94514 Sw:O21276 Tr:Q9SGT7 Tr:O29158 Sw:O12460
                   Tr:O80401 Tr:O65334 Tr:O65335 Tr:O9U5W4"
                   /codon_start=1
                   /pseudo
                   /evidence=not_experimental
repeat_region      34666. .34977
                   /note="AluSg repeat: matches 11. .297 of consensus"
repeat_region      36947. .36978
                   /note="16 copies 2 mer aa 87% conserved"
misc_feature       38608. .39160
                   /note="match: GSS: Em:AQ353745"
polyA_site         complement(40339)
                   /gene="ELAVL2"
mRNA               complement(join(40339. .43119,43683. .43721,51615. .51840,
                   55153. .55306,81257. .81360,112241. .112481,171869. .171992))
                   /gene="ELAVL2"
                   /note="match: cDNAs: Em:U29088 Em:X85111 Em:M62843
                   Em:U17602 Em:U12431 Em:L26405 Em:U17599 Em:U29148
                   Em:U17597 Em:D26158 Em:S83320 Em:D31953 Em:AF176675
                   match: ESTs: Em:AA545382 Em:AA081584 Em:AI843112
                   Em:AA656788 Em:AA692890 Em:AI145457 Em:AV117363
                   Em:AW524152 Em:AA416369 Em:AI028574 Em:AW491988
                   Em:AA549705 Em:D61361 Em:AV362071 Em:AI837658 Em:AV094529


Query Match         3.9%;     Score 89.2;   DB 9;   Length 183483;
Best Local Similarity  64.6%;     Pred. No. 9.9e-14;
Matches 133;   Conservative    0;   Mismatches 73;   Indels    0;   Gaps    0;

Qy    891 ctggaccaggtgggcaaactcttcctgtaaagaaccagacaggaactcatttaggctc 950
          || || ||||| | |||||| |||||| |||||  || ||||||||
Db  10360 CTAGCACAGAGGTCAGCAAACTTTCTCTGTAAAGGACCAGGATCATGATAAATATTTTAGGTTT 10419

Qy    951 tgtgtgccatatggtctcagtcacaactactcatctctgccctctgtagcacgaaagcaat 1010  .
          |||| |||||||| ||||||| |||| | |||||| |||| | || |  | |||||| .
Db  10420 TGTGGGCCATATGGTCTCTGTCACAACTACTCCACTCTGCTGTGTTCTAGGGCAAAAGCAGT 10479

Qy   1011 tagcaacaatatgtcaacaaacatatgtgaccccatgaaaactttattattatggatac 1070
          || |||| || ||| || ||| || ||  | |||| ||||||| || ||||||||| ||
Db  10480 TATAGAAAATACATAAATGAATGAGCCGTGTTCCATAAAAGTATTTATGGACAC 10539

Qy   1071 ggaaactgaaataatgtcttcctt 1096
          |||| |||| ||||| || ||
Db  10540 TGAAATTTGAATTTCATGTAATTTTT 10565


Search completed: June  6, 2002, 12:16:57
Job time: 14004 sec
```

```
                  GenCore version 4.5
          Copyright (C) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        June  6, 2002, 09:51:13 ; Search time 360.11 Seconds
                                          (without alignments)
                                          10899.076 Million cell updates/sec

Title:          US-09-515-369B-1
Perfect score:  2286
Sequence:       1 taatacgactcactaggg.........tgacttccacggctgggacg 2286
Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0
Searched:       1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:       3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N_Geneseq_032802:*
          1:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
          2:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
          3:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
          4:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
          5:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
          6:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
          7:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
          8:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
          9:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
         10:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
         11:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
         12:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
         13:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
         14:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
         15:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
         16:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
         17:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
         18:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
         19:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
         20:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
         21:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
         22:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
         23:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
         24:   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Match Length	DB	ID	Description
	1	2286	100.0	2286	22	AAH26595	Human melanoma dif
c	2	81.8	3.6	37437	22	AAK84165	Human immune/haema
c	3	81.4	3.6	37442	22	AAK84166	Human immune/haema
c	4	81	3.5	5294	22	AAH14600	Human cDNA sequenc
c	5	77.6	3.4	26591	22	AAL36313	Human musculoskele
	6	75.8	3.3	481	22	AAK74761	Human immune/haema
	7	73	3.2	1837	22	ABA04430	Human PP1345 prote
c	8	71.6	3.1	628	24	AAH98619	Human EST-derived
c	9	71.4	3.1	286	21	AAA41353	Human secreted exp

c	10	71.2	3.1	407	21	AAC27788	Human secreted pro
c	11	70	3.1	123219	23	AAH88703	Human DNA sequence
c	12	69.8	3.1	50000	22	AAX23517	Human kidney amino
c	13	69.4	3.0	8029	22	AAH99747	Human protein enco
c	14	69.4	3.0	46765	22	AAS99306	DNA encoding Aldeh
c	15	69.4	3.0	47319	22	AAK64813	Human immune/haema
c	16	69.4	3.0	47319	22	AAK72230	Human immune/haema
	17	69.4	3.0	236303	22	AAS11614	Human genomic DNA
	18	68.6	3.0	325791	22	AAS43104	Human Oestrogen re
	19	68.2	3.0	97662	22	AAF83908	Genomic sequence o
	20	68	3.0	31140	22	ABA08065	Human ovarian and
c	21	68	3.0	31140	22	AAL06791	Human reproductive
c	22	67.8	3.0	17700	22	AAK68945	Human immune/haema
c	23	67.6	3.0	2515	22	AAD17444	Human granulocyte
	24	67.6	3.0	6235	21	AAZ29159	Human G-CSF genom1
	25	67.6	3.0	6679	22	AAZ29170	Partial sequence o
	26	67.6	3.0	6679	22	AAD17443	Human granulocyte
	27	66.6	2.9	289	22	AAS39009	Novel human diagno
	28	66.4	2.9	466	22	AAH33971	Human colon cancer
c	29	66.4	2.9	655	22	AAL17197	Human breast cance
	30	66	2.9	871	22	AAH04280	Human cDNA clone (
	31	66	2.9	1712	22	AAH13903	Human cDNA sequenc
	32	66	2.9	2291	22	AAH17522	Human cDNA sequenc
	33	66	2.9	7496	22	AAK70490	Human immune/haema
	34	66	2.9	7496	22	AAK74614	Human immune/haema
c	35	66	2.9	10225	22	AAI99021	Human excretory re
c	36	66	2.9	10225	22	AAI63371	Human kidney relat
c	37	66	2.9	24292	22	ABA15939	Human nervous syst
	38	66	2.9	28731	22	AAK64887	Human immune/haema.
	39	66	2.9	28731	22	AAK69606	Human immune/haema
	40	66	2.9	28731	22	AAK70491	Human immune/haema
	41	66	2.9	28731	22	AAK74615	Human immune/haema
c	42	66	2.9	28731	22	AAK84942	Human immune/haema
	43	65.6	2.9	872	22	AAH98871	Human EST-derived
c	44	65.4	2.9	519	22	AAL35469	Human musculoskele
c	45	64.8	2.8	15044	22	AAS36290	Human cardiovascul

ALIGNMENTS

```
RESULT
  1
AAH26595
ID   AAH26595 standard; DNA; 2286 BP.
XX
AC   AAH26595;
XX
DT   12-NOV-2001  (first entry)
XX
DE   Human melanoma differentiation associated gene-7 (first entry)
DE
XX
KW   Melanoma differentiation associated gene-7; Mda-7; promoter; human;
KW   neuroblastoma; astrocytoma; glioblastoma multiforme;
KW   cervical cancer; breast cancer; colon cancer; prostate cancer;
KW   osteosarcoma; chondrosarcoma; tumour; therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   promoter        1..2240
FT                   /*tag= a
FT   protein_bind    138..146
FT                   /*tag= b
FT                   /bound_moiety= "C/EBP"
FT   protein_bind    446..454
FT                   /*tag= c
FT                   /bound_moiety= "C/EBP"
FT   protein_bind    773..777
FT                   /*tag= d
FT                   /bound_moiety= "C/EBP"
FT   protein_bind    801..807
FT                   /*tag= e
FT                   /bound_moiety= "AP-1"
FT
```

Human melanoma differentiation associated gene-7 (Mda-7) promoter.

```
FT   protein_bind   1109..1113
FT                  /tag= f
FT                  /bound_moiety= "C/EBP"
FT   protein_bind   1207..1214
FT                  /tag= g
FT   protein_bind   1209..1214
FT                  /bound_moiety= "AP1"
FT                  /tag= h
FT   protein_bind   1530..1537
FT                  /bound_moiety= "C/EBP"
FT                  /tag= i
FT   protein_bind   2115..2121
FT                  /bound_moiety= "AP1"
FT                  /tag= j
FT   protein_bind   2126..2133
FT                  /bound_moiety= "AP1"
FT                  /tag= k
FT   TATA_signal    2210..2214
FT                  /bound_moiety= "C/EBP"
FT                  /tag= l
XX
PN   WO200164921-A1.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06782.
XX
PR   29-FEB-2000; 2000US-0515369.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Fisher PB, Madireddi MT;
XX
DR   WPI; 2001-565508/63.
XX
PT   Melanoma differentiation associated gene-7 promoter capable of
PT   treating cancer comprises directing transcription of heterologous
PT   coding sequence encoding tumour suppressor polypeptide positioned
PT   downstream, useful for treating cancer.
XX
PS   Claim 1; Fig 6A-B; 132pp; English.
XX
CC   The present sequence is that of the promoter region of the human
CC   melanoma differentiation associated gene-7 (Mda-7). It includes
CC   2240 bp from the 5' flanking region of the Mda-7 gene. The
CC   promoter was isolated from a human placental genomic library
CC   using a PCR-based method. The Mda-7 promoter exhibits melanocyte
CC   tissue specificity and can only be activated in the targeted
CC   tissue, i.e. the skin. Therefore, a gene of interest driven by
CC   the Mda-7 promoter will be differentially expressed in these cells,
CC   minimising systemic toxicity. A recombinant expression construct
CC   in which the human Mda-7 promoter is operably linked to a coding
CC   sequence encoding a tumour suppressor protein is claimed. The
CC   tumour suppressor is preferably p21, retinoblastoma protein or p53.
CC   A host cell comprising the expression construct is also claimed,
CC   and is preferably a tumour cell selected from a melanoma,
CC   neuroblastoma, astrocytoma, glioblastoma multiforme, cervical
CC   cancer, breast cancer, lung cancer or prostate cancer cell.
CC   The promoter preferably comprises nucleotides 1-2240 of the present
CC   sequence. A pharmaceutical composition including the recombinant
CC   expression construct is used in a claimed method of treating
CC   melanoma, neuroblastoma, astrocytoma, glioblastoma multiforme,
CC   cervical cancer, breast cancer, colon cancer, prostate cancer,
CC   osteosarcoma, chondrosarcoma or a cancer of the central nervous
CC   system.
XX
SQ   Sequence 2286 BP; 547 A; 592 C; 510 G; 637 T; 0 other;
```

```
Qy          1 taatagactcactatagggcgtcgactcgatcacctttgaacccaggctgcctgcct   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 taatagactcactatagggcgtcgactcgatcacctttgaacccaggctgcctgcct   60

Qy         61 ccaagctgtactcatcatagattctcaactgatgtgggccaaggttctaggttct  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ccaagctgtactcatcatagattctcaactgatgtgggccaaggttctaggttct  120

Qy        121 ctcctgaccttccttctgaagtaataatgctatgtatgataagctcatcggaggctgagcc  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ctcctgaccttccttctgaagtaataatgctatgtatgataagctcatcggaggctgagcc  180

Qy        181 aggcacatgtttgctgaactatccatgttatatattcctccagacagagtgagct  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 aggcacatgtttgctgaactatccatgttatatattcctccagacagagtgagct  240

Qy        241 actcacgatcccaggtgtaccctgaggccagccaaggtgtatccatgacctcatgccctct  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 actcacgatcccaggtgtaccctgaggccagccaaggtgtatccatgacctcatgccctct  300

Qy        301 gttccagcctgccctttaacagctcatccccgcctgccccccaccctgcctatctgcaga  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 gttccagcctgccctttaacagctcatccccgcctgccccccaccctgcctatctgcaga  360

Qy        361 cagtagtctaggattcagctgcctgccctgggggctcattctcctctcagcttcctgctta  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 cagtagtctaggattcagctgcctgccctgggggctcattctcctctcagcttcctgctta  420

Qy        421 gctgtctcctgcctcccactactactccagcactctcacctggtcttcctttctccg  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 gctgtctcctgcctcccactactactccagcactctcacctggtcttcctttctccg  480

Qy        481 tctcatcactgcctcttgacatcttatctcatagtagttagggttcttggtaat  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 tctcatcactgcctcttgacatcttatctcatagtagttagggttcttggtaat  540

Qy        541 gccctaaatccacatggtgggaagggggagtgggggaagagagtgcgctgtggggctgt  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 gccctaaatccacatggtgggaagggggagtgggggaagagagtgcgctgtggggctgt  600

Qy        601 gcctacttctggaggtaagactccgggcctccaggaacaaaggattcaggctggtggtgca  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 gcctacttctggaggtaagactccgggcctccaggaacaaaggattcaggctggtggtgca  660

Qy        661 gctatagccaagcagactgctggccagggattgcaaaggagtatattgttgttgcttaagaa  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 gctatagccaagcagactgctggccagggattgcaaaggagtatattgttgttgcttaagaa  720

Qy        721 aataaacaacactgagtgatgagaggagggggtgttggtgcagagagattgggaa  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 aataaacaacactgagtgatgagaggagggggtgttggtgcagagagattgggaa  780

Qy        781 gagtctgccaaggtgtgttcactactcctcctctttcttcatcccactgagctgg  840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 gagtctgccaaggtgtgttcactactcctcctctttcttcatcccactgagctgg  840

Qy        841 aggcagttatcctgtccccacgtcacattccgttccacctgatatttaggctctgtgtgccat  900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 aggcagttatcctgtccccacgtcacattccgttccacctgatatttaggctctgtgtgccat  900

Qy        901 gttgggcaaactcttcctgtaaagaaccagacaggaactatcttaggctgtgctgtgccat  960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 gttgggcaaactcttcctgtaaagaaccagacaggaactatcttaggctgtgctgtgccat  960

Qy        961 atggtctcagtcacaacatatgtgcaccgctctgtagcacgaaagcaattagcaacaat 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 atggtctcagtcacaacatatgtgcaccgctctgtagcacgaaagcaattagcaacaat 1020

Qy       1021 atgtcaacaaacatatgtgaccccatgaaaacttattatggataccggaaacctga 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 atgtcaacaaacatatgtgaccccatgaaaacttattatggataccggaaacctga 1080

Qy       1081 aaataatgtcttttctttgattttccccaatcattaaaaaacgtaaaaactactctta 1140
```

Query Match   100.08;   Score 2286;   DB 22;   Length 2286;
Best Local Similarity   100.08;   Pred. No. 0;
Matches 2286;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Db   1081  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           aaaatgtcttttcttttgattttcccaatccatcattaaaaaacgtaaaaacctactctta  1140

Qy   1141  ggtcgcaaggttaagccattccagctgttcagtgcagtggcagctgatttggcttgtgacct  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  ggtcgcaaggttaagccattccagctgttcagtgcagtggcagctgatttggcttgtgacct  1200

Qy   1201  acagttggccaatccctgattcctcagggatgtgggcaaatacttat  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  acagttggccaatccctgattcctcagggatgtgggcaaatacttat  1260

Qy   1261  gggaagtgctgattaaaacagagttaagaagcatcagacatttccaggacgcggctagcac  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  gggaagtgctgattaaaacagagttaagaagcatcagacatttccaggacgcggctagcac  1320

Qy   1321  atgccaggggctcctaactgacctcattggattcatctgttcatggaggatcttgcaag  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  atgccaggggctcctaactgacctcattggattcatctgttcatggaggatcttgcaag  1380

Qy   1381  acaagaattcctcaaacctagagtctgaggactgtgctgggaaacactgctctgcttg  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  acaagaattcctcaaacctagagtctgagaactgtgctgggaaacactgctctgcttg  1440

Qy   1441  atgccctactgggcacatggcacatggttcagcctgcttgctagctctggagataggg  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  atgccctactgggcacatggcacatggttcagcctgcttgctagctctggagataggg  1500

Qy   1501  gtccagagctcttgactgctgccctggcagtcttgacacatcacgctgtctgtcccctgagt  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  gtccagagctcttgactgctgccctggcagtcttgacacatcacgctgtctgtcccctgagt  1560

Qy   1561  ggttcagagccaccaccaggccaagactagccccagactagcccaccaccagagggctcctttctg  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  ggttcagagccaccaccaggccaagactagccccagactagcccaccaccagagggctcctttctg  1620

Qy   1621  ggcttgtccatgtgctacattttccttattctccctggttccagaaacctaaggagaggcac  1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  ggcttgtccatgtgctacattttccttattctccctggttccagaaacctaaggagaggcac  1680

Qy   1681  atttggttgagtgattataacctaggaccatgggtagctgcatgtcaggaaaactc  1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  atttggttgagtgattataacctaggaccatgggtagctgcatgtcaggaaaacactc  1740

Qy   1741  ctcaacttcctggcctgatgattaaggaggtacttacaggttattcttcgctgt  1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  ctcaacttcctggcctgatgattaaggaggtacttacaggttattcttcgctgt  1800

Qy   1801  ggactactgtccagcatgaatagggcatcattatttgaatattttgacaggaaggagac  1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  ggactactgtccagcatgaatagggcatcattatttgaatattttgacaggaaggagac  1860

Qy   1861  tgtgtatgctgcacagtaatgtatttacatgtgacagagttaccaagcacctct  1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  tgtgtatgctgcacagtaatgtatttacatgtgacagagttaccaagcacctct  1920

Qy   1921  gtgttgttttgccttgtttattacacttggacaaattttaaaatttatacatgcag  1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  gtgttgttttgccttgtttattacacttggacaaattttaaaatttatacatgcag  1980

Qy   1981  agactgcagcgcgagaagctaagagactgccccctgcccacacagccagtggtagagcc  2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  agactgcagcgcgagaagctaagagactgccccctgcccacacagccagtggtagagcc  2040

Qy   2041  tgaactcaaaccaggtctcatctcacctcagggctgcttcccatcgctgtattgtc  2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  tgaactcaaaccaggtctcatctcacctcagggctgcttcccatcgctgtattgtc  2100

Qy   2101  cttaaagtgatgggtgactagccagcaatgaagtaattctctaggaaaagcatgaccaatttcc  2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101  cttaaagtgatgggtgactagccagcaatgaagtaattctctaggaaaagcatgaccaatttcc  2160

Qy   2161  cttctctccacctcctcttttcctccaccctccccatcagcccccatatatgccc  2220
           |||||||||||||||||
```

```
Db   2161  cttctctccacctcctcttttcctccaccctccccatcagcccccatatatgccc  2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2221  aaatctccacaaagcctgttgctgcttgctgcttgctgcgaaacctttacttctgaaatgacct  2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221  aaatctccacaaagcctgttgctgcttgctgcgcaaacctttacttctgaaatgacttccacggct  2280

Qy   2281  gggacg  2286
           ||||||
Db   2281  gggacg  2286

RESULT  2
AAK84165/C
ID     AAK84165  standard; DNA; 37437 BP.
XX
AC     AAK84165;
XX
DT     07-NOV-2001  (first entry)
XX
DE     Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38977.
XX
KW     Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW     cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS     Homo sapiens.
XX
PN     WO200157182-A2.
XX
PD     09-AUG-2001.
XX
PF     17-JAN-2001; 2001WO-US01354.
XX
PR     31-JAN-2000; 2000US-0179065.
PR     04-FEB-2000; 2000US-0180628.
PR     24-FEB-2000; 2000US-0184664.
PR     02-MAR-2000; 2000US-0186350.
PR     16-MAR-2000; 2000US-0189874.
PR     17-MAR-2000; 2000US-0190076.
PR     18-APR-2000; 2000US-0198123.
PR     19-MAY-2000; 2000US-0205515.
PR     07-JUN-2000; 2000US-0209467.
PR     28-JUN-2000; 2000US-0214886.
PR     30-JUN-2000; 2000US-0215135.
PR     07-JUL-2000; 2000US-0216647.
PR     07-JUL-2000; 2000US-0216880.
PR     11-JUL-2000; 2000US-0217487.
PR     11-JUL-2000; 2000US-0217496.
PR     14-JUL-2000; 2000US-0218290.
PR     26-JUL-2000; 2000US-0220963.
PR     26-JUL-2000; 2000US-0220964.
PR     14-AUG-2000; 2000US-0224518.
PR     14-AUG-2000; 2000US-0224519.
PR     14-AUG-2000; 2000US-0225213.
PR     14-AUG-2000; 2000US-0225214.
PR     14-AUG-2000; 2000US-0225266.
PR     14-AUG-2000; 2000US-0225267.
PR     14-AUG-2000; 2000US-0225268.
PR     14-AUG-2000; 2000US-0225270.
PR     14-AUG-2000; 2000US-0225447.
PR     14-AUG-2000; 2000US-0225757.
PR     14-AUG-2000; 2000US-0225758.
PR     14-AUG-2000; 2000US-0225759.
PR     18-AUG-2000; 2000US-0226279.
PR     22-AUG-2000; 2000US-0226681.
PR     22-AUG-2000; 2000US-0226868.
PR     22-AUG-2000; 2000US-0227182.
PR     23-AUG-2000; 2000US-0227009.
PR     30-AUG-2000; 2000US-0228924.
PR     01-SEP-2000; 2000US-0229287.
PR     01-SEP-2000; 2000US-0229343.
PR     01-SEP-2000; 2000US-0229344.
PR     01-SEP-2000; 2000US-0229345.
PR     05-SEP-2000; 2000US-0229509.
```

```
PR   05-SEP-2000;  2000US-0229513.
PR   06-SEP-2000;  2000US-0230437.
PR   06-SEP-2000;  2000US-0230438.
PR   08-SEP-2000;  2000US-0231242.
PR   08-SEP-2000;  2000US-0231243.
PR   08-SEP-2000;  2000US-0231244.
PR   08-SEP-2000;  2000US-0231413.
PR   08-SEP-2000;  2000US-0231414.
PR   08-SEP-2000;  2000US-0232080.
PR   08-SEP-2000;  2000US-0232081.
PR   12-SEP-2000;  2000US-0231968.
PR   14-SEP-2000;  2000US-0232397.
PR   14-SEP-2000;  2000US-0232398.
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.
PR   14-SEP-2000;  2000US-0233065.
PR   21-SEP-2000;  2000US-0234223.
PR   21-SEP-2000;  2000US-0234274.
PR   25-SEP-2000;  2000US-0234997.
PR   25-SEP-2000;  2000US-0234998.
PR   26-SEP-2000;  2000US-0235484.
PR   27-SEP-2000;  2000US-0235834.
PR   27-SEP-2000;  2000US-0235836.
PR   29-SEP-2000;  2000US-0236327.
PR   29-SEP-2000;  2000US-0236367.
PR   29-SEP-2000;  2000US-0236368.
PR   29-SEP-2000;  2000US-0236369.
PR   29-SEP-2000;  2000US-0236370.
PR   02-OCT-2000;  2000US-0236802.
PR   02-OCT-2000;  2000US-0237037.
PR   02-OCT-2000;  2000US-0237038.
PR   02-OCT-2000;  2000US-0237039.
PR   02-OCT-2000;  2000US-0237040.
PR   13-OCT-2000;  2000US-0239935.
PR   13-OCT-2000;  2000US-0239937.
PR   20-OCT-2000;  2000US-0240960.
PR   20-OCT-2000;  2000US-0241221.
PR   20-OCT-2000;  2000US-0241785.
PR   20-OCT-2000;  2000US-0241786.
PR   20-OCT-2000;  2000US-0241787.
PR   20-OCT-2000;  2000US-0241808.
PR   20-OCT-2000;  2000US-0241809.
PR   20-OCT-2000;  2000US-0241826.
PR   01-NOV-2000;  2000US-0244617.
PR   08-NOV-2000;  2000US-0246474.
PR   08-NOV-2000;  2000US-0246475.
PR   08-NOV-2000;  2000US-0246476.
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249297.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251988.
PR   05-DEC-2000;  2000US-0256719.
PR   06-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI;  2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure;  SEQ ID NO 38977;  3071pp + Sequence Listing;  English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  .
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production.(I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome   .
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 37437 BP; 10357 A; 6715 C; 6866 G; 13499 T; 0 other;

     Query Match             3.6%;  Score 81.8;  DB 22;  Length 37437;
     Best Local Similarity   62.9%;  Pred. No. 1.7e-12;
     Matches  144;  Conservative   0;  Mismatches  82;  Indels   3;  Gaps   1;

Qy        807 actctcctttttcttcatccactgagctgagcagtatcctgtccccacgtca 866
              |||||||| | ||| ||||||||||  || || | |||  |  |||| || |
Db      32330 AGTCCACTTCATTCTTTCATGTCTACTTCTTTCCAGTATATATTCCTAAAAACAACATAA 32271

Qy        867 ca--ttcctactcccgttcccatgcctggaccaggttgggcaaactcttcctgtaaa 923
              || |||| || ||||||||||||||| |||||||||||||||| ||| ||| |||
Db      32270 AAATTGTCCTTTGCAATGGTCAGATATCTAACACAGGGTCAGCAAACTTTTTCTGAAAA 32211

Qy        924 gaaccagcaggaaactatttaggctctgtgtgccatatggtctcagtcaacaactactca 983
              | |||| | || |||||||||| |||| ||| ||||| |||||| ||| ||| ||||||
Db      32210 GAGCCAGATGGTAAATATTTAGGCTTTGTGGGCCAGACATTCTCTGTTGCAACTATTCA 32151

Qy        984 tctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaac 1032
```

```
Db  32150  TTTCTGCTGCTATAGTACAAAAGCAGCTATAAACAATAAGTAAGCAAAC 32102

RESULT 3
AAK84166/c
ID   AAK84166 standard; DNA; 37442 BP.
XX
AC   AAK84166;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38978.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
```

```
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251988.
PR   05-DEC-2000;  2000US-0256719.
PR   06-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 38978; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 37442 BP; 10361 A; 6715 C; 6860 G; 13506 T; 0 other;

     Query Match           3.6%;   Score 81.4;  DB 22;  Length 37442;
     Best Local Similarity  63.0%;           Pred. No. 2.2e-12;
     Matches 143;    Conservative   0;  Mismatches  81;  Indels   3;  Gaps   3;

Qy     809 tctcctctttttttcatctccactgagctggaggcagttatcctgtccccacgtcaca 868
           ||  || ||||||||||| ||| |||   ||||| ||||  | |||    || |  |
Db   32329 TCCACTTCATTCTTTCATGTCACTTCTTCTTCCAGTATATAT ATTCCTAAAAACACATAAAA 32270

Qy     869 ---ttcctactccgttcccatgctgggaccaggttgggcaaactcttcctgtaaaga 925
           |||   ||     ||||||||| | |||   |||||||||  |||  |||| |||
Db   32269 ATTGTCCTTTGCAATGGTCAGATATCTAACACAGGGGTCAGCAAACTTTTCTGAAAAGA 32210

Qy     926 accagacaggaactattttaggctcgtgtgccatatggtgccatcagtctcagtacaactactcatc 985
           ||||| ||||| |||| | ||||||| |||  |||| | ||||||||| || |||| ||||
Db   32209 GCCAGATGGTAAATATTTTAGGCTTTGTGGGCCAGCATTCTCTGTTGCAACTATTCATT 32150

Qy     986 tctgtccctgtagcacgaagcaattagcaacaacatatgtcaacaaac 1032
           ||||  ||  ||| ||   |||||| || ||||| |||| || | ||||
Db   32149 TCTGCTGCCTATAGTACAAAAAGCAGCTATAAACAATAAGTAAGCAAAC 32103
```

```
RESULT   4
AAH14600/c
ID   AAH14600 standard; cDNA; 5294 BP.
XX
AC   AAH14600;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12215.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000;  2000EP-0116126.
XX
PR   29-JUL-1999;  99JP-0248036.
PR   27-AUG-1999;  99JP-0300253.
PR   11-JAN-2000;  2000JP-0118776.
PR   02-MAY-2000;  2000JP-0183767.
PR   09-JUN-2000;  2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 12215; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 5294 BP; 1637 A; 878 C; 1099 G; 1680 T; 0 other;

     Query Match           3.5%;   Score 81;  DB 22;  Length 5294;
     Best Local Similarity 74.5%;           Pred. No. 9.5e-13;
     Matches 102;  Conservative   0;  Mismatches  35;  Indels   0;  Gaps   0;

Qy     894 gacccaggttggcaaactcttcctgtaaagaaccaggagaccaggagaactatttaggctcgt 953
           ||  |||| |||||||||||||| ||||| | | |||||||| ||| |||||| |||||||
Db    1915 GACCAGGAATCAGAAAACTTTTTCTGTAAAGGCCAGAGAGTAAGTATTTTAGGTTTTGT 1856

Qy     954 gtgccatatggtctcagtcagtcacaactactcatctctgcctctgtagcacgaaagcaattag 1013
           | ||||||||||||| | ||| || |||| |||||||| | ||||||| ||||||||||||||
```

```
Db    1855 GGGCCATATGGTCTCTGCTGGCAATTACTTAACTCTGCCATCATAGCACGAAAGCAGCTAC 1796
            |||||||| ||||||||| |||
Qy    1014 caacaatatgtcaacaa 1030
Db    1795 AGACAATATGTACGCAA 1779

RESULT 5
AAL36313/c
ID   AAL36313 standard; DNA; 26591 BP.
XX
AC   AAL36313;
XX
DT   08-JAN-2002 (first entry)
XX
DE   Human musculoskeletal system related polynucleotide SEQ ID NO 2678.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein;
KW   musculoskeletal system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155367-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01338.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
```

```
PR    17-NOV-2000;  2000US-0249217.
PR    17-NOV-2000;  2000US-0249218.
PR    17-NOV-2000;  2000US-0249244.
PR    17-NOV-2000;  2000US-0249245.
PR    17-NOV-2000;  2000US-0249264.
PR    17-NOV-2000;  2000US-0249265.
PR    17-NOV-2000;  2000US-0249297.
PR    17-NOV-2000;  2000US-0249299.
PR    17-NOV-2000;  2000US-0249300.
PR    01-DEC-2000;  2000US-0250160.
PR    01-DEC-2000;  2000US-0250391.
PR    05-DEC-2000;  2000US-0251030.
PR    05-DEC-2000;  2000US-0251988.
PR    05-DEC-2000;  2000US-0256719.
PR    06-DEC-2000;  2000US-0251479.
PR    08-DEC-2000;  2000US-0251856.
PR    08-DEC-2000;  2000US-0251868.
PR    08-DEC-2000;  2000US-0251869.
PR    08-DEC-2000;  2000US-0251989.
PR    08-DEC-2000;  2000US-0251990.
PR    11-DEC-2000;  2000US-0254097.
PR    05-JAN-2001;  2001US-0259678.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Barash SC,  Ruben SM;
XX
DR    WPI; 2001-451937/48.
XX
PT    Isolated polypeptide for treating, preventing and/ or prognosing
PT    disorders related to the musculoskeletal system including
PT    musculoskeletal cancers and also for testing and detection e.g.
PT    diagnosis -
XX
PS    Example 2; SEQ ID NO 2678; 781pp + Sequence Listing; English.
XX
CC    The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC    (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC    for preventing, treating or ameliorating medical conditions e.g. by
CC    protein or gene therapy. The genes are isolated from a range of human
CC    tissues disclosed in the specification. The nucleic acids, proteins,
CC    antibodies and (ant)agonists are useful in the diagnosis, treatment
CC    and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC    other cancers of the adrenal gland, bone, bone marrow, breast,
CC    gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC    disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC    anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC    multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC    (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC    healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC    and (f) infectious diseases such as viral, bacterial, fungal and
CC    parasitic infections.
CC    Note: The sequence data for this patent did not form part of the
CC    printed specification, but was obtained in electronic format directly
CC    from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence 26591 BP; 7489 A; 6354 C; 5953 G; 6795 T; 0 other;
```

```
Query Match     3.4%;  Score 77.6;  DB 22;  Length 26591;
Best Local Similarity   70.3%;    Pred. No. 2.4e-11;
Matches 104;    Conservative    0;  Mismatches  44;  Indels    0;  Gaps     0;

Qy     894 gacccaggtgggcaaactcttcctgtaaagaaccagacaggaactcattttaggctctgt 953
           || |   |  || ||||| |  ||||  | | |  |   |  ||| | | | |||||||
Db   23936 GAGCAGGGGTTAGCAAAATTTTCTGTCTTAGAGTCATATAGTAAATATTTTAAACTCTGC 23877

Qy     954 gtgccatatggtctcagtcacaactactcatctctgcctctgtagcacgaaagcaattag 1013
           || ||||||| ||| | |||||| |  ||||||||  ||||||| ||| | |||||||
Db   23876 AGGTTCCTGTGGTCTCTGTCTCTCAGCTACTCAACTCTGCCATTGTAGCCATGGAAGCAATTGC 23817

Qy    1014 caacaatatgtcaacaaacatatgtgac 1041
           | |||||||| |||||| ||||| |
```

```
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0255484.
PR   27-SEP-2000;   2000US-0255834.
PR   27-SEP-2000;   2000US-0255836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.

PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 29573; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 481 BP; 140 A; 103 C; 79 G; 159 T; 0 other;

Query Match             3.3%;  Score 75.8;  DB 22;  Length 481;
Best Local Similarity   68.9%;
Matches  104;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      912 tctcctgtaaagaccacaggaactcatttaggctccgtgtgccatatggctcagt 971
                | || |||||| | ||||||||| ||||||||| |||||||| | ||||||| ||
Db      282 tttttttgtaaagggccagacagcaggaaatatttaggcttttaggcttgtgggtcgtatggtctctgt 341

Qy      972 cacaactactcatctctgcctctgtagcacgaaagcaattagcacgacaatatgtcaacaa 1031
                ||||| |||| | |||| |||||| || | ||||| | ||| |||||
Db      342 tgcaactgctcaagtttgccactgtagaactaaagcagccgtggatgatacctaaacaa 401

Qy     1032 catatgtgaccccatgaaaactttattttatt 1062
                | ||||| | || ||||||  |||||
Db      402 cgggtgtgactgtgttccaaaaaatttatt 432


RESULT 7
ABA04430
ID   ABA04430 standard; cDNA; 1837 BP.
XX
AC   ABA04430;
XX
DT   11-MAR-2002 (first entry)
XX
```

DE   Human PP1345 protein encoding cDNA SEQ ID NO:19/21.
XX
KW   Human; PP1345; cancer suppression; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   174..566
FT   /*tag= a
FT   /product= "PP1345"
XX
PN   CN1313315-A.
XX
PD   19-SEP-2001.
XX
PF   13-MAR-2000; 2000CN-0111989.
XX
PR   13-MAR-2000; 2000CN-0111989.
XX
PA   (SHAN-) SHANGHAI INST ONCOLOGY.
XX
PI   Gu J, Yang S;
XX
DR   WPI; 2002-042193/06.
DR   P-PSDB; ABB04713.
XX
PT   New human protein able to suppress growth of cancer cells and its
PT   encoding polynucleotide sequence -
XX
PS   Claim 5; Page 33 (Disclosure); 42pp; Chinese.
XX
CC   The present sequence encodes human PP1345 protein, which has cancer-
CC   suppressing activity. The present invention also describes a method
CC   for the preparation of the protein by recombination, and the application
CC   of the protein in treating diseases such as cancer.
XX
SQ   Sequence 1837 BP; 478 A; 471 C; 407 G; 481 T; 0 other;

Query Match   3.2%; Score 73; DB 24; Length 1837;
Best Local Similarity 64.5%; Pred. No. 1.2e-10;
Matches 109; Conservative   0; Mismatches 60; Indels 0; Gaps 0;

Qy   918 tgtaaagaaccagacagaactatttaggtctgtgtgtcacatagtctcagtcacaac 977
          | |||| |||| || |||||| || ||||||| ||| |||||||| |||||||
Db   1639 tttaaaggaacagatagtaaatatttaagcttgtgagccacatggtctctgtcacaac 1698

Qy   978 tactcatcctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatatg 1037
          || /|| |||||| |||||| |||| || ||||||| || |||||| || ||
Db   1699 tattcagctctgctattgtaatacaaaagcagctggacgtatgtaaatgaatgaatg 1758

Qy   1038 tgacccctgaaaacttattattatggatacggaaacctgaaaataa 1086
          ||| || | || ||||| |||||| |||||| |||| || ||
Db   1759 tggctgtgttccataaaacttattcccaaaaaaaaaaaaaaaaaa 1807

RESULT 8
AAH98619/c
ID   AAH98619 standard; cDNA; 628 BP.
XX
AC   AAH98619;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Human EST-derived coding sequence SEQ ID NO: 476.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Homo sapiens.
XX

PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI   Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM23960.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Claim 1; Page 514; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention.
XX
SQ   Sequence 628 BP; 190 A; 122 C; 131 G; 185 T; 0 other;

Query Match   3.1%; Score 71.6; DB 22; Length 628;
Best Local Similarity 60.9%; Pred. No. 1.7e-10;
Matches 151; Conservative   0; Mismatches 94; Indels 3; Gaps 2;

Qy   894 gacccaggttggcaaactcttcctgtaaagaaccagacagaactatttaggtctgt 953
          |||| |||| |||| || ||||| |||||| ||| |||||| ||| ||||||| -
Db   452 GACCAGGGTTCAGCAAGCTTTTCTGTAAAAGGCCAAATAGTAAATATTTAGGCTTTGT 393

Qy   954 gtgccatatggtctcagtcacaactactcatcctgcctctgtagcacgaaagcaattag 1013
          - ||||||||||| || ||| ||| |||||||| || || ||||| |
Db   392 GGACCCATATGGTCTC--TGCTAACTACTCAATTCTGCCATTGAAGTGCAAAAGCAACCAT 335

Qy   1014 caacaatatgtcaacaaacatatgtgaccccatgaaaactttatttattatggatacgga 1073
          | |||||| || ||||| ||| || || |||| ||||| || |||||||
Db   334 GTA-AATGAATGTACGTGGCTGTGTTTATTTGTGGACACTGAAATTTAAATTTCATAATT 276

Qy   1074 aacctgaaaataatgtccttctttttgatttttcccaatcattaaaaaacgtaaaaact 1133
          |||| || ||| || ||| || ||||||||||| ||||||||||| -
Db   275 TTCACATAAAATATTTCTTTTTGATTTTTTTTTCCCAACAATTAAAAAATTTTAAGCC 216

Qy   1134 actcttag 1141
          | |||||
Db   215 ATGCTTAG 208

RESULT 9
AAA41353/c
ID   AAA41353 standard; cDNA; 286 BP.
XX
AC   AAA41353;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:93.
XX
KW   Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;

KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO20021990-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999;   99WO-US24205.
XX
PR 15-OCT-1998;   98US-0104435.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX
DR WPI; 2000-317937/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 198; 618pp; English.
XX
CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 286 BP; 88 A; 63 C; 50 G; 85 T; 0 other;

Query Match   3.1%;   Score 71.4; DB 21; Length 286;
Best Local Similarity 72.1%;   Pred. No. 1.2e-10;
Matches 93; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 899 aggttgggcaaactcttcctgtaaagaaccagacagggaactcatttaggctcgtgtgcc 958
        |||||||||||||||| |||||||||||| || ||||||||| |||||| |||||||||
Db 149 AGCCAGATAAACATTTTTCTGTAAAGAGTCAGACAGTAAATATTTTAGGCCTGGTCGTCC 90

Qy 959 atatggtctcagtcacaactactcatctctgcctctgtagcacgaaagcaaattagcaaca 1018
        |||||||||| ||| ||| || ||||||||||||||| || | || | | ||| |||

Db 89 ATATGGTCTCTGCTGCGCCAACTACACAACTCTGCGTTCTTCTGTGTAAAAGCAGCAATAGACA 30

Qy 1019 atatgtcaa 1027
         ||||||| ||
Db 29 ATATGTAAA 21

RESULT 10
AAC27788/c
ID AAC27788 standard; cDNA; 407 BP.
XX
AC AAC27788;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 31863.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999;   99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 31863; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 407 BP; 118 A; 72 C; 84 G; 127 T; 6 other;

Query Match   3.1%;   Score 71.2; DB 21; Length 407;
Best Local Similarity 63.6%;   Pred. No. 1.7e-10;
Matches 124; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

Qy 887 atgcctggaccaggtgggcaaactcttcctgtaaagaaccagacaggaactcatttag 946
        |||| || |||| |||||||||||| || |||||||| ||||||||||||||| |||
Db 402 ATGCCTAGACAACAGACTGGCAAAATTTTCTGTAAAGAGCCAGACAGTAAATATGTTAG 343

Qy 947 gctctgtgtgccatatggtctcagtcacaactactcatctctgcctctgtagcacgaaag 1006
        ||| ||| || ||| | |||| | | |||| || |||||| |||||||| ||||||
Db 342 GCTTTGCAGGCTACATGATCTCTGTTGCAACTACTCAACTCTGCCCTCACAATGCAAAAG 283

Qy 1007 caattagcaacaacatatgtcaacaaacatatgtgaccccatgaaaacttattattatgg 1066
         ||| | | | | ||||| ||| | |||| | |||| ||| |||| |||||

Db   282 CAGCCATTGAGAATGCTCAAACCAATGAGTGTGGCTGTGTTNCAA--TAACATTTTATGG 225
          || ||| |||| || |||| || || || | | ||||||||

Qy   1067 atacggaaacctgaa 1081
          | | |||| ||||

Db   224 ACATTGAAATTTGAA 210

RESULT 11
AAH88703
ID   AAH88703 standard; DNA; 123219 BP.
XX
AC   AAH88703;
XX
DT   26-FEB-2002 (first entry)
XX
DE   Human DNA sequence SEQ ID 543.
XX
KW   Single nucleotide polymorphism; SNP; biallelic marker; human;
KW   central nervous system disorder; CNS; ds.
XX
OS   Homo sapiens.
XX
PN   WO200151659-A2.
XX
PD   19-JUL-2001.
XX
PF   11-JAN-2001; 2001WO-IB00116.
XX
PR   13-JAN-2000; 2000US-0175854.
XX
PA   (GEST) GENSET.
XX
PI   Chu T, Blumenfeld M, Cohen D;
XX
DR   WPI; 2001-483085/52.
XX
PT   Isolated polynucleotides, useful for genotyping nucleic acids for
PT   biallelic markers for the diagnosis of depression, comprises central
PT   nervous system disorder related biallelic marker -
XX
PS   Disclosure; Page 439-472; 519pp; English.
XX
CC   The present invention relates to biallelic markers derived from human
CC   genes involved in central nervous system (CNS) disorders (see
CC   AAH88161-AAH88702). The markers have a single nucleotide polymorphism
CC   (SNP) and are useful in determining the genetic predisposition of
CC   individuals to CNS disorders, by identifying the nucleotides at a set of
CC   genetic markers in a biological sample, where the markers comprise at
CC   least one CNS disorder related marker. The present sequence was used
CC   to illustrate the invention.
XX
SQ   Sequence 123219 BP; 27921 A; 32327 C; 32359 G; 30437 T; 165 other;

Query Match   3.1%; Score 70; DB 23; Length 123219;
Best Local Similarity 61.5%; Pred. No. 1e-08;
Matches 112; Conservative   0; Mismatches 70; Indels   0; Gaps   0;

Qy   895 accaggttgggcaaactcttcctgtaaagacagcaggaactatttaggctctgtg 954
          | | |||||| || ||||||||| |||| ||| | || ||| || ||||| ||||
Db 17375 agcagggtcagcaaacttttctggaaaggtagatagacaactgtcatggctttgtg 17434

Qy   955 tgccatatggtcttcagtcacaacactctctgcctctgagcacgaaagcaattagc 1014
          ||||| || ||||| | ||| || || | ||| || ||| |||| |||| |||||
Db 17435 ggccacgtggcctccatcgcgtcattcaactctatgctgtgagcaggaaagcagccata 17494

Qy   1015 aacaatagtcaacaacatatgtgaccccatgaaaacttt attattattggatacggaa 1074
          ||||||||| ||||| || ||| |||| ||||||| || ||||| |||||||||
Db 17495 gacaatagtcacaagggccgtgctgcatgccaataaaacttt atttattataaaacat 17554

Qy   1075 ac 1076
          ||

Db 17555 ac 17556

RESULT 12
AAX23517/c
ID   AAX23517 standard; DNA; 50000 BP.
XX
AC   AAX23517;
XX
DT   23-JUN-1999 (first entry)
XX
DE   Human kidney aminopeptidase P genomic DNA fragment 1.
XX
KW   Aminopeptidase; human; AmP; gene therapy; treatment; AmP-deficiency;
KW   prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;
KW   arterial stenosis; industrial protein feed; malabsorption syndrome;
KW   proteinaceous waste degradation; additive; immunohistochemistry; ss.
XX
OS   Homo sapiens.
XX
PN   WO9911799-A2.
XX
PD   11-MAR-1999.
XX
PF   02-SEP-1998; 98WO-US18426.
XX
PR   02-SEP-1997; 97US-0057854.
XX
PA   (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.
XX
PI   Ryan JW, Sprinkle TJC, Venema RC;
XX
DR   WPI; 1999-205193/17.
XX
PT   Nucleic acid encoding human aminopeptidase P
XX
PS   Claim 13; Page 80-109; 201pp; English.
XX
CC   This invention describes the isolation of a novel human aminopeptidase P
CC   (AmP). This protein is used to produce recombinant AmP and can be used
CC   for gene therapy for treating AmP-deficiency conditions. Its fragments
CC   are used as primers and probes to identify patients with homozygous and
CC   heterozygous AmP deficiency, including prenatal diagnosis (patients
CC   defective in AmP are at risk of developing angioedema if treated with
CC   angiotensin-converting enzyme inhibitors), also as antisense inhibitors
CC   in cases of excessive AmP expression. The product of the invention is
CC   also used to identify AmP-expressing sequences in other animals and to
CC   generate transgenic animals, and comparisons of genomic sequences are
CC   used to detect mutations. AmP inhibitors are potentially useful as
CC   antihypertensive agents and to prevent or treat arterial (re)stenosis
CC   or atherosclerosis. The structure of AmP is used to design synthetic
CC   substrates, e.g. for use in AmP assays. AmP, which hydrolyzes N-terminal
CC   imido bonds, can be used to degrade industrial protein feeds to free
CC   amino acids, to degrade proteinaceous wastes, as additives in enzyme
CC   formulations used to treat malabsorption syndrome and for studying its
CC   biological role. Antibodies against AmP are used in immunohistochemical
CC   methods to study AmP distribution.
XX
SQ   Sequence 50000 BP; 13187 A; 12125 C; 11842 G; 12846 T; 0 other;

Query Match   3.1%; Score 69.8; DB 20; Length 50000;
Best Local Similarity 72.7%; Pred. No. 7e-09;
Matches 104; Conservative   0; Mismatches 37; Indels   2; Gaps   1;

Qy   889 gcctggaccaggttggcaaactcttcctgtaaagaaccagcaggaactatttaggc 948
          | || | |||||| ||||| |||||||||||| |||||| |||| || |||||||
Db   8967 GTCTAGATCAAAGTTTGCCAAA--CTTTCTGTAAAGGGTCAGATAGTAAGTATTTTAGAC 8910

Qy   949 tctgtgtgccatatggtcttcagtcacaacactctctgcctctgtagcacgaaagca 1008
          ||||||| ||||||| | |||| || ||| || | | || | || || || ||
Db   8909 TCTGCAGACCATACAGTCTCTGTAGCAACCACTCAACTCTGCTGTTCGCAGCCCCAAAGCA 8850

```
Qy   1009 attagcaacaatatgtcaacaaa 1031
          | || ||||||| | |||||
Db   8849 AGTATAGACAATATATAAACAA 8827
```

CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders.
XX
SQ   Sequence 8029 BP; 2064 A; 1942 C; 1905 G; 2118 T; 0 other;

RESULT 13
AAH99747
ID   AAH99747 standard; cDNA; 8029 BP.
XX
AC   AAH99747;
XX
DT   16-OCT-2001 (first entry)
XX
DE   Human protein encoding cDNA sequence SEQ ID NO:582.
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO20153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35017.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   P-PSDB; AAM25806.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS   Claim 1; Page 622-624; 1217pp; English.
XX
CC   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and
CC   cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production. The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation,
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, buros, ulcers,
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic

```
Qy    889 gctggaccaggttggcaaactcttcctgtaaagaacagcaggaactatttaggc 948
          ||||||| ||||| |||||||||| |||||||||| |||| |||||  ||||||
Db    682 gctagatcaggagctggcaaactttcttaaggagacaagtagtaaatcttttaggc 741

Qy    949 tctgtgtgccatatgtctcagtcacacactcatcctgcctctgtagcac------ 1001
          | || ||| || ||| || |||| ||| |||||| ||||||||  |||
Db    742 ttgcaggctatactgccctgtcccaactattcaactctgcaagaaagcagtcccaga 801

Qy   1002 ------------gaaagcaattagcaacaacatatgtcaacaaacatatgtgaccccatga 1048
                         ||||||||| || ||  |||| ||||||||||
Db    802 caataccataatgaatggtgtggctatgctctaataacatttgtttatggacactga 861

Qy   1049 aaacttattt-attatggatacggaaacctgaaaataaatgtcttctttttgattttttc 1107
          |||| |||| || || ||| |||||||| |||| ||||||||||||||||
Db    862 aatctgaatttcatagaatttcaggtgtcatgaatattgttcctctttgattttctg 921

Qy   1108 cccaatcattaaaaacgtaaaaactactcttaggtcgcaaggttaagccattctcagct 1167
          |||||||||| ||||||  |||| ||| | ||||||||||||  ||  |||
Db    922 cccaatcatttgaaacatataaaccattcttagctcacagg--------ttgtataaa 972

Qy   1168 tagcagtggcaggctgggattggcttgtgacctacagttggccaatccctgat 1220
          - |||||||||| ||||||||||| | ||||||||||||||||||| |||||||
Db    973 aacaggtggcaggccaggtttgattcgtgggccatagtttgccaaccccctgat 1025
```

RESULT 14
AAS99306/c
ID   AAS99306 standard; DNA; 46765 BP.
XX
AC   AAS99306;
XX
DT   12-MAR-2002 (first entry)
XX
DE   DNA encoding Aldehyde dehydrogenase 5 family, member A1 (ALDH5A1).
XX
KW   Aldehyde dehydrogenase 5 family member A1; ALDH5A1;
KW   succinate-semialdehyde dehydrogenase; gene therapy; probe;
KW   antisense technology; allele specific oligonucleotide; ASO;
KW   4-hydroxybutyric aciduria; metabolic disease; transgenic animal;
KW   chromosome 6p22; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   variation   replace(3749,T)
FT   /*tag= a
FT   variation   /standard_name= "Single nucleotide polymorphism"
FT   replace(3800,G)
FT   /*tag= b
FT   variation   /standard_name= "Single nucleotide polymorphism"
FT   replace(3829,T)
FT   /*tag= c
FT   variation   /standard_name= "Single nucleotide polymorphism"
FT   replace(3875,G)
FT   /*tag= d
FT   variation   /standard_name= "Single nucleotide polymorphism"
FT   replace(3876,A)
FT   /*tag= e
FT   variation   /standard_name= "Single nucleotide polymorphism"
FT   replace(3941,T)
FT   /*tag= f
FT   /standard_name= "Single nucleotide polymorphism"

Query Match   3.0%; Score 69.4; DB 22; Length 8029;
Best Local Similarity 57.2%; Pred. No. 3.2e-09;
Matches 202; Conservative   0; Mismatches 121; Indels 30; Gaps   3;

```
FT   exon            4050..4403
FT                   /tag= g
FT                   /number= 1
FT   CDS             4050..42765
FT                   /tag= h
FT                   /product= "ALDH5A1"
FT                   /note= "Aldehyde dehydrogenase 5 family member A1.
FT                   Specifically claimed in claim 25"
FT   variation       replace(4059,G)
FT                   /tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4111,A)
FT                   /tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4155,C)
FT                   /tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4179,G)
FT                   /tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          4404..11575
FT                   /tag= m
FT                   /number= 1
FT   variation       replace(4430,G)
FT                   /tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            11576..11659
FT                   /tag= o
FT                   /number= 2
FT   intron          11660..12315
FT                   /tag= p
FT                   /number= 2
FT   exon            12316..12486
FT                   /tag= q
FT                   /number= 3
FT   variation       replace(12415,T)
FT                   /tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12422,T)
FT                   /tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          12487..13921
FT                   /tag= t
FT                   /number= 3
FT   variation       replace(12559,T)
FT                   /tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13855,G)
FT                   /tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13917,T)
FT                   /tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            13922..14038
FT                   /tag= x
FT                   /number= 4
FT   variation       replace(14021,T)
FT                   /tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          14039..24219
FT                   /tag= z
FT                   /number= 4
FT   variation       replace(14046,A)
FT                   /tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14179,A)
FT                   /tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            24220..24363
FT                   /tag= ac
FT                   /number= 5
FT   intron          24364..29453
FT                   /tag= ad

FT                   /number= 5
FT   intron          24598..31819
FT                   /tag= ae
FT                   /number= 6
FT   variation       replace(29397,T)
FT                   /tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            29454..29597
FT                   /tag= ag
FT                   /number= 6
FT   variation       replace(29469,A)
FT                   /tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            31820..31978
FT                   /tag= ai
FT                   /number= 7
FT   intron          31979..37049
FT                   /tag= aj
FT                   /number= 7
FT   exon            37050..37219
FT                   /tag= ak
FT                   /number= 8
FT   intron          37220..41171
FT                   /tag= al
FT                   /number= 8
FT   exon            41172..41230
FT                   /tag= am
FT                   /number= 9
FT   variation       replace(41217,C)
FT                   /tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          41231..42559
FT                   /tag= ao
FT                   /number= 9
FT   variation       replace(41237,A)
FT                   /tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(41403,C)
FT                   /tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(42380,T)
FT                   /tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            42560..42765
FT                   /tag= as
FT                   /number= 10
FT
XX
PN   WO200190119-A2.
XX
PD   29-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-US16558.
XX
PR   19-MAY-2000; 2000US-205849P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kliem SE, Koshy B, Tanguay DA;
XX
DR   WPI; 2002-089912/12.
DR   P-PSDB; AAU73594.
XX
PT   New genetic variants of human aldehyde dehydrogenase 5 family, member
PT   A1, ALDH5A1 gene for treating metabolic diseases and for expressing
PT   ALDH5A1 protein useful in identifying drugs to treat 4-hydroxybutyric
PT   aciduria -
XX
PS   Disclosure: Fig 1; 151pp; English.
XX
CC   The invention describes an isolated polynucleotide comprising a
CC   nucleotide sequence which is a polymorphic variant of a reference
CC   sequence for the aldehyde dehydrogenase 5 family, member A1
CC   (succinate-semialdehyde dehydrogenase) (ALDH5A1) gene or its fragment.
```

```
CC   The polypeptide is useful for screening for drugs targeting it by
CC   contacting the ALDH5A1 polymorphic variant with a candidate agent and
CC   assaying for binding activity. The polypeptide and haplotypes are useful
CC   for identifying an association between a trait such as a clinical
CC   response to a drug targeting ALDH5A1 and a haplotype ALDH5A1 gene.
CC   Transgenic animals are also useful for studying expression of the ALDH5A1
CC   isogenes in vivo, for in vivo screening and testing of drugs against
CC   ALDH5A1 protein and for testing the efficacy of therapeutic agents and
CC   compounds for 4-hydroxybutyric aciduria and metabolic diseases in a
CC   biological system. Antibodies are useful for diagnostic and prognostic
CC   formats and therapeutic methods, for immunoprecipitating the polypeptide
CC   from solution, for detecting ALDH5A1 protein isoforms in biological
CC   samples, frozen tissue sections, for use in immunocytochemical,
CC   immunohistochemical and immunofluorescence techniques. The polynucleotide
CC   is useful for gene therapy and antisense gene therapy. This sequence
CC   encodes ALDH5A1 (located on chromosome 6p22), and forms a reference
CC   sequence on which polymorphic sequences encoding isoforms of the ALDH5A1
CC   protein are based, described in the method of the invention.
XX

Query Match          3.0%;   Score 69.4;   DB 24;   Length 46765;
Best Local Similarity  57.2%;   Pred. No. 8.8e-09;
Matches  202;  Conservative   0;  Mismatches 121;  Indels   30;  Gaps   3;

Qy    889 gcctggaccaggttgggcaaactcttcctgtaaagaaccagacaggaactattttaggc 948
              ||| || ||  |||||||| || ||| ||| ||| || ||| | ||||||||
Db  45589 GTCTAGATCAGGAGCTGGCAAACTTTTTTTCTTAAGGAGACAAGTAGTAAATCTTTTAGGC 45530

Qy    949 tctgtgtgccatatggtctcagtcacaaactactcatctctgcctctgtagcac------- 1001
              || |||| ||||||  |||| ||| | |||| ||| |||||| ||| |||
Db  45529 TTTGCAGGCTATACTGTCCCTGTCCCCAACTATTCAACTCTGCCAAGAAAGCAGTCCCAGA 45470

Qy   1002 --------gaaagcaattagcaacaatatgtcaacaaacatatgtgacccccatga 1048
                      || || | ||| |  ||||
Db  45469 CAATACCATAAATGAATGGGTGTGGCTATGCTCTAATAACATTTTGTTTATGGACACTGA 45410

Qy   1049 aaacttattt-attatgatacggaaacctgaaaataatgtcttttcttttgatttttc 1107
              |||| |||| ||  || |  || ||| |||||||| |  |||||||||||||
Db  45409 AATCTGAATTTCATAGAATTTCAGGTGTCATGAAATATATTGTTCCTTCTTTTGATTTTCTG 45350

Qy   1108 cccaatcattaaaaacgtaaaaactactcttaggtcgcaaggttaagccattccagct 1167
              |||||||| |||||||||| | |||| ||||| ||||| | |||| ||||| || ||
Db  45349 CCCAATCATTTGAAAACATATAAAACCATTCTTAGCTCACAGG-------TTGTATAAA 45299

Qy   1168 tagcagtggcaggctggattggcttgacctacagttggccaatccctgat 1220
              |||||||| | |  |||| || ||| ||| ||| ||||||| |||||||
Db  45298 AACAGGTGGCAGGCCAGGTTTGATTCGTGGGCCCATAGTTTGCCACCCCTGAT 45246
```

RESULT 15
AAK64813/c
ID   AAK64813 standard; DNA; 47319 BP.
XX
AC   AAK64813;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19625.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.

PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216680.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.

CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 47319 BP; 13708 A; 8453 C; 9720 G; 15438 T; 0 other;

Query Match
Best Local Similarity   3.0%;   Score 69.4; DB 22; Length 47319;
                         66.2%;   Pred. No. 8.9e-09;
Matches 100; Conservative   0; Mismatches 51; Indels   0; Gaps   0:

Qy   881 tttcccatgcctggacccaggttggccaaactcttcctgtaaagaaccagacaggaacta 940
          |||||||||| ||| ||| |||||||||||| || ||||| ||| ||||| ||| |
Db 18549 TTTACTAAGTCTAAATCAGAGATTGGCAAACTTTTTCTGTCAAGAGCCAGGCATTAATTA 18490

Qy   941 tttagctctgtgtgccatatggtctcagtcacaactactcatctctgcctctgagca 1000
          |||||||| ||| ||| ||| ||||||| | ||| ||| || ||| | ||||
Db 18489 TTTTAGGCTTTGTAAGCCACACAGTCTCTACGGCAACCACTCAACTCTGTCCATATAGCC 18430

Qy   1001 cgaaagcaattagcaacaatatgtcaacaaa 1031
          | ||||| |||| || || |||| ||| ||
Db 18429 CCAAAGCTGCTACAGATGATATATAAATGAA 18399

Search completed: June 6, 2002, 12:05:15
Job time: 8042 sec

us-09-515-369b-1.rng

Thu Jun 6 16:21:04 2002

```
          GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   June 6, 2002, 08:14:48 ; Search time 2404.01 Seconds
                                    (without alignments)
                                    12834.406 Million cell updates/sec

Title:          US-09-515-369B-1
Perfect score:  2286
Sequence:       1 taatagactcactataggg.........tgacttccacggctgggacg 2286

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters:          27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
         1:  em_estba:*
         2:  em_esthum:*
         3:  em_estin:*
         4:  em_estmu:*
         5:  em_estov:*
         6:  em_estpl:*
         7:  em_estro:*
         8:  em_htc:*
         9:  gb_est1:*
        10:  gb_est2:*
        11:  gb_htc:*
        12:  gb_gss:*
        13:  em_gss_hum:*
        14:  em_gss_inv:*
        15:  em_gss_pln:*
        16:  em_gss_vrt:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	81	3.5	480	10	BI262402	BI262402 602953836
2	80.6	3.5	354	10	T58770	T58770 yb80h02.s1
3	80.4	3.5	390	10	L44398	L44398 HUMEST1E3 H
4	78.2	3.4	521	12	AQ695956	AQ695956 HS_2160_A
5	78.2	3.4	435	9	AW183569	AW183569 xj76h04.x
6	78.2	3.4	459	9	AI125442	AI125442 qd89b02.x
7	78.2	3.4	628	12	AQ378792	AQ378792 RPCI11-15
c 8	77.4	3.4	393	9	AI128823	AI128823 qa94d06.s
9	77.4	3.4	411	10	W76494	W76494 zd67d03.r1
c 10	77.4	3.4	680	12	AG154552	AG154552 Pan trogl
11	77.2	3.4	400	9	AI926274	AI926274 wo43g06.x
c 12	77	3.4	322	9	AI240516	AI240516 qh12h02.x
13	77	3.4	386	9	AA443938	AA443938 zv51c03.s
14	77	3.4	393	9	AA44117	AA44117 zv51c03.r
c 15	77	3.4	417	9	AI208768	AI208768 qg38c10.x
c 16	76.8	3.4	509	10	BI061153	BI061153 IL3-UT011
c 17	76.6	3.4	327	12	B98848	B98848 CIT-HSP-228
c 18	76.2	3.3	499	12	AQ277351	AQ277351 CITBI-E1-
c 19	76.2	3.3	644	12	AQ039307	AQ039307 CIT-HSP-2
c 20	76.2	3.3	5529	12	AF101969	AF101969 AF101969
c 21	76	3.3	573	12	AQ633901	AQ633901 RPCI-11-4
22	75.6	3.3	530	12	AQ508112	AQ508112 RPCI-11-2
23	75.6	3.3	628	12	AQ020407	AQ020407 CIT-HSP-2
24	75.2	3.3	799	10	BG535645	BG535645 60256353
c 25	74.6	3.3	488	9	AV604853	AV604853 AV604853
c 26	73.6	3.2	348	12	AQ035111	AQ035111 CIT-HSP-2
c 27	73.6	3.2	640	12	AG093736	AG093736 Pan trogl
c 28	73.2	3.2	381	9	AW574982	AW574982 UI-HF-BN0
29	73.2	3.2	512	12	AQ768179	AQ768179 HS_3160_B
30	73	3.2	519	12	AQ029411	AQ029411 RPCI11-39
c 31	72.8	3.2	302	9	AI559884	AI559884 th87a12.x
32	72.2	3.2	427	10	AQ815705	AQ815705 HS_5394_A
33	72	3.2	312	10	BG230513	BG230513 naj44a06.
34	72	3.1	416	9	AW835797	AW835797 QV4-LT001
c 35	72	3.1	764	10	BI915216	BI915216 60184745
c 36	71.4	3.1	2032	11	AF161549	AF161549 Homo sapi
37	71.2	3.1	794	10	BF983586	BF983586 60230575
c 38	71.2	3.1	327	9	AI344682	AI344682 qp09h04.x
c 39	71.2	3.1	363	9	AI358986	AI358986 qy19h02.x
c 40	71.2	3.1	372	9	AI370636	AI370636 ta38g12.x
c 41	71.2	3.1	380	9	AI244892	AI244892 qj98f05.x
c 42	71.2	3.1	380	9	AI298789	AI298789 qm91c02.x
c 43	71.2	3.1	388	9	AI291781	AI291781 qm75b05.x
c 44	71.2	3.1	393	9	AI214952	AI214952 qm31d12.x
c 45	71.2	3.1	394	9	AI439755	AI439755 ti61a07.x

ALIGNMENTS

```
RESULT 1
BI262402
LOCUS       BI262402       480 bp    mRNA    linear    EST 17-JUL-2001
DEFINITION  602953836T1 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5080033 3',
            mRNA sequence.
ACCESSION   BI262402
VERSION     BI262402.1   GI:14822582
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 480)
  AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue procurement: Dr. Louis Staudt, M.D., Ph.D. cDNA Library
            Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1846  row: g   column: 10
            High quality sequence start: 27
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1..480
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5080033"
                     /clone_lib="NIH_MGC_99"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size selected >500bp for average insert size
```

```
                 1.8kb. Library constructed by Ling Hong in the laboratory
                 of Gerald M. Rubin (University of California, Berkeley)
                 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                 II RT (Life Technologies). Note: this is a NIH_MGC
                 Library."

BASE COUNT      142 a    101 c    112 g    125 t
ORIGIN

Query Match            3.5%;     Score 81;   DB 10;  Length 480;
Best Local Similarity 74.5%;     Pred. No. 7e-10;
Matches 102;  Conservative 0;  Mismatches 35;  Indels 0;  Gaps 0;

Qy   894 gacccaggttgggcaaactcttcctgtaaagaaccagacaggaactatttaggctctgt 953
         | ||||||||||| ||| || ||||||||||||||||||||||| |||  |||||||||
Db   156 GACCAGGAATCAGAAAACTTTTTCTGTAAAGGGCCAGAGTAAGTAATATTTTAGGTTTTGT 215

Qy   954 gtgccatatggtctcagtcacaactactcatctcgccctctgtagcacgaaagcaattag 1013
         | |||||||||| ||| || |||||| || ||| ||||||||| |||||||||| ||
Db   216 GGGCCATATGGTCTCTGTGGCAATTACTTAACTCTGCCATCATAGCACGAAAGCAGCTAC 275

Qy   1014 caacaatatgtcaacaa 1030
          |||||||| |||
Db   276 AGACAATATGTACGCAA 292


RESULT  2
T58770
LOCUS       T58770         354 bp   mRNA     linear   EST 09-FEB-1995
DEFINITION  yb80h02.s1 Stratagene liver (#937224) Homo sapiens cDNA clone
            IMAGE:77523 3', mRNA sequence.
ACCESSION   T58770
VERSION     T58770.1  GI:660607
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 354)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,N., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohling,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE       Generation and analysis of 280,000 human expressed sequence tags
JOURNAL     Genome Res. 6 (9), 807-828 (1996)
MEDLINE     97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 2225
            High quality sequence stops: 271  Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2225    Std Error: 0.00
            Seq primer: -21ml3
            High quality sequence stop: 271.
FEATURES             Location/Qualifiers
     source          1..354
                     /organism="Homo sapiens"
                     /db_xref="GDB:497268"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:77523"
                     /clone_lib="Stratagene liver (#937224)"
                     /sex="male"
                     /dev_stage="49 years old"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: liver; Vector: pBluescript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
```

```
                 dT. Hepatectomy from normal male caucasian. Average insert
                 size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                 GAATTCGGCACGAG 3'; -3' adaptor sequence: 5'
                 CTCGAGTTTTTTTTTTTTTTTTTT 3'"

BASE COUNT      105 a     69 c     77 g     99 t      4 others
ORIGIN

Query Match            3.5%;     Score 80.6;  DB 10;  Length 354;
Best Local Similarity 74.3%;     Pred. No. 8.1e-10;
Matches 101;  Conservative 0;  Mismatches 35;  Indels 0;  Gaps 0;

Qy   894 gacccaggttgggcaaactcttcctgtaaagaaccagacaggaactatttaggctctgt 953
         | ||||||||||| ||| || ||||||||||||||||||||||| |||  |||||||||
Db   181 GACCAGGGATCAGAAAACTTTTTCTGTAAAGGGCCCNGAGTAAGTAATATTTTAGGTTTTGT 240

Qy   954 gtgccatatggtctcagtcacaactactcatctcgccctctgtagcacgaaagcaattag 1013
         | |||||||||| ||| || |||||| || ||| ||||||||| |||||||||| ||
Db   241 GGGCCATATGGTCTCTGTGGCAATTACTTAACTCTGCCATCATAGCACGAAAGCAGCTAC 300

Qy   1014 caacaatatgtcaaca 1029
          |||||||| ||
Db   301 AGACAATATGTACGCAA 316


RESULT  3
L44398
LOCUS       L44398         390 bp   mRNA     linear   EST 17-JAN-1996
DEFINITION  HUMEST1E3 Human thymus NSTH II Homo sapiens cDNA, mRNA sequence.
ACCESSION   L44398
VERSION     L44398.1  GI:1048762
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 390)
AUTHORS     Lamerdin,J.E., Athwal,R.S., Patanjali,S., Weissman,S. and Carrano
            ,A.V.
TITLE       Chromosomal localization and expressed sequence tag generation of
            clones from a normalized human adult thymus cDNA library
JOURNAL     Genome Res. 5 (4), 359-367 (1995)
MEDLINE     96356441
COMMENT     Contact: Human Genome Center, Biology and Biotechnology Research
            Program
            Lawrence Livermore National Laboratory
            Livermore, CA 94550, USA.
FEATURES             Location/Qualifiers
     source          1..390
                     /organism="Homo sapiens"
                     /db_xref="GDB:D19S801E"
                     /db_xref="taxon:9606"
                     /clone_lib="Human thymus NSTH II"
                     /note="From adult human thymus NSTH II cells; randomly
                     primed; normalized"

BASE COUNT      118 a     85 c     64 g     123 t
ORIGIN

Query Match            3.5%;     Score 80.4;  DB 10;  Length 390;
Best Local Similarity 62.4%;     Pred. No. 9.3e-10;
Matches 126;  Conservative 0;  Mismatches 76;  Indels 0;  Gaps 0;

Qy   890 cctggacccaggttgggcaaactcttcctgtaaagaaccagacaggaactatttaggct 949
         |||| ||||||| ||| ||||| |||||||||||||||| |||||||||||||||| ||
Db   29  CCTAGATCAAGAGTAGGAAAACTTTTTCGTCAAGAGCAAGAAGATAGTAACTATTTGGTCT 88

Qy   950 ctgtgtgccatatggtctcagtcacaactactcatctcgccctctgtagcacgaaagcaa 1009
         |||  |||| |||| |||| ||| || ||||| ||| | |||| |||||| |||||| ||
Db   89  TTGTTAGCCATATGGTCCCCTGCCAATTACTCAACTTTGCCACTGTAGTGAGAAAGCCA 148

Qy   1010 ttagcaacaatatgtcaacaaacatatgtgacccccatgaaaacttatttatttatttggata 1069
          || || ||| || | |  ||| | | |||
```

```
Db   149  CCAGAGACAATACATAAGCAAGGGAGCTAGCTATCGTTCCAATAGAACTTTATTTATAAA  208

Qy  1070  cggaaacctgaaaataatgtct  1091
          |||  |||| || ||
Db   209  AACAAATGGTCAACTATAGTTT  230
```

RESULT 4
AQ695956
```
LOCUS       AQ695956    521 bp    DNA     linear   GSS 06-JUL-1999
DEFINITION  HS_2160_A2_F03_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate-2160 Col=6 Row=K, DNA sequence.
ACCESSION   AQ695956
VERSION     AQ695956.1  GI:5386204
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web server: http://www.htsc.washington.edu
            Plate:2160  row: K  column: 6
            Seq primer: T7
            Class: BAC ends
FEATURES             Location/Qualifiers
     source          1..521
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate-2160 Col=6 Row=K"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT   150 a   128 c   92 g   148 t
ORIGIN

Query Match             3.4%;
Best Local Similarity  73.5%;   Score 78.4;  DB 12;  Length 521;
Matches 100;  Conservative  0;  Mismatches 36;  Indels 0;  Gaps 0;  Pred. No. 3.3e-09;

Qy   906  gcaaactccttcctgtaaagaaccaggagaactattttaggctctgtgtgccatatggt  965
          |||||| || ||||||||||  ||||| |  |||| ||||| || || ||||||||
Db   290  GCAAACTTTTCTGTAAAGGGCCAGATAGTAAATATTGTAGACTTTGCAGGACATATGTT  349

Qy   966  ctccagtcacaactactcatcctgcctctgtagcacgaaagcaattagcaacaatatgtc  1025
          || | |  |||||||  | |||||| |  ||||| || |   |||||  ||||||| ||
Db   350  CTCCTGTTCCAACTACTTAGCTCTGCGGTTGAGCCACCAAAAGCAGCCATAGACAAATGCA  409

Qy  1026  aacaaacatatgtgac  1041
          ||||| ||||||
Db   410  GACAAATGCATGTGAC  425
```

RESULT 5
AW183569
```
LOCUS       AW183569    435 bp    mRNA    linear   EST 18-NOV-1999
DEFINITION  xj76h04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2663191 3', mRNA sequence.
ACCESSION   AW183569
VERSION     AW183569.1  GI:6452083
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 435)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1..435
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2663191"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHI19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT   149 a   82 c   60 g   144 t
ORIGIN

Query Match             3.4%;
Best Local Similarity  59.6%;   Score 78.2;  DB 9;  Length 435;
Matches 201;  Conservative  0;  Mismatches 108;  Indels 28;  Gaps 3;

Qy   901  gttgggcaaactccttcctgtaaagaaccaggacaggagaactatttaggctctgtgtgccat  960
          | ||||| || ||||| || |||||| |||| || | |||||| ||  |||||||||||
Db   108  GGTCAGCGAAACTTTTCTGTAAAGGGCCATACAATACATATTTCAGTCTCTGCAGACCAT  167

Qy   961  atggtcagtcacaactactcatcctgcctctgtagcacgaaagcaattagcaacaat  1020
          |||| | |||| |||||||||  | |||| |||||||  ||||| ||||||  ||||
Db   168  TTGGTCTTTGCTGAAACTACTCAACTCGTCAACATAACAAAGCTGCCACAGACAAT  227

Qy  1021  atgtcacaaacatatgtgacccatgaaacttatttat---tatggatacggaaacc  1077
          || | |   ||||  |||| |||| ||  | |||||||   |||||||  |||||
Db   228  ATGTAAACAAATGAGGGTGACTGTAATAAAATGTTATTTATTGAACAACTGAAATTTGAACT  287

Qy  1078  tgaaaataatgtctttct------------tttgattttcccaatcattaaa  1121
          | ||| | |||| | |||            |||||||| | |||||| |||||
Db   288  TCAAATCATTTTCATGCTTCCAAAACATTCTACTTTTGACTATTTTCAACCATTTAAA  347

Qy  1122  aacgtaaaaactactcttaggtcgcaaggttaagccattctcagcttagcagtggcaggc  1181
          || ||||||  |||| |||| |  ||| || ||||||| | ||| ||||| ||  ||
Db   348  AAAGTAAAAACCATTCTTAGGTGACAGCGTGTAGAAAAACT--------ACTCATAGGC  398

Qy  1182  tggatttggctgtgacctacagttggccaatccctg  1218
          |  | ||||||||||| |||||||||||||| |||
Db   399  TTAATCTGCCCACAAGCCATAGTTCGCCAATGCCTG  435
```

RESULT 6

```
AI125442
LOCUS       AI125442               459 bp    mRNA    linear   EST 28-OCT-1998
DEFINITION  qd89b02.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1736619
            3', mRNA sequence.
ACCESSION   AI125442
VERSION     AI125442.1  GI:3593956
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 459)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham.
            Insert Length: 592    Std Error: 0.00
FEATURES             Location/Qualifiers
     source          1..459
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1736619"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAGCGGCCGCGCCAATTTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT    159 a     88 c     61 g    151 t
ORIGIN

Query Match        3.4%;    Score 78.2;   DB 9;   Length 459;
Best Local Similarity  59.1%;  Pred. No. 3.6e-09;
Matches  204;  Conservative   0;  Mismatches 113;  Indels  28;  Gaps  3;

Qy   901 gttgggcaaactcttcctgtaaagaaccagacaggaactatttaggctctgtgtgccat 960
         || ||||||| || |||||||||   ||| ||  ||||| ||| |||||  | ||||||||
Db   105 GGTCAGCAAACTTTTTCTGTAAAGGGCCATACATATTCAGTCTCTGCAGACCAT 164

Qy   961 atggtctcagtcacaactactcatctgcctctgtagcacgaaagcaattagcaacaat 1020
         ||||| |  |||| |||||||||  ||||||| |||| |||||||| ||| | |||||
Db   165 TTGGTCTTTGCGAAACTACTCAACTCTGCTAACATAACAAAAGCTGCCACAGACAAT 224

Qy   1021 atgtcaacaaacatatgtgaccccatgaaaacttatttat---tatggatacggaaacc 1077
          |||| |||||| |  |||||||||  || |||  || ||||   ||||||||  |||||
Db   225 ATGTAAACAAATGAGGTGACTGTAATAAAARATGTTATTTATGAACACTGAAATTTGAACT 284

Qy   1078 tgaaaataatgtctttct----------------ttgattttccccatcattaaaa 1121
          | ||| |||| | ||||                || ||||| | |||| || ||||
Db   285 TCAAATCATTTTCATGCTTCCCAAACATTCTACTTTTGACTATTTTCAACCATTTAA 344

Qy   1122 aacgtaaaaactactcttaggtcgcaaggttaagccattctcagctagcagtggcagc 1181
          |  |||||||| || |||||||   |||| ||| |||||||||||||||  | ||||
Db   345 AAGTAAAAACCATTCTTAGGTGACAGGCTGTAGAAAACT--------AGTCATAGGC 395
```

```
Qy   1182 tggatttggcttgtgacctacagttggccaatccctgatcccaa 1226
          || ||||| |||| ||||||||||| |||||||||  ||||| ||
Db    396 TTAATCTGGCCACAAGCCATAGTTCGCCAATGCCGTCACACTA 440

RESULT 7
AQ378792/c
LOCUS       AQ378792               628 bp    DNA    linear   GSS 20-MAY-1999
DEFINITION  RPC11-151B24.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-151B24,
            DNA sequence.
ACCESSION   AQ378792
VERSION     AQ378792.1  GI:4349815
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 628)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
TITLE       Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1997)
            Other_GSSs: RPCI1-151B24.TV
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@djong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html .
            Seq primer: SP6
            Class: BAC ends
FEATURES             Location/Qualifiers
     source          1..628
                     /organism="Homo sapiens"
                     /db_xref="GDB:7557647"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-151B24"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"         1 others
BASE COUNT    173 a    133 c    136 g    185 t
ORIGIN

Query Match        3.4%;    Score 78.2;   DB 12;   Length 628;
Best Local Similarity  72.1%;  Pred. No. 4e-09;
Matches  101;  Conservative   0;  Mismatches  39;  Indels   0;  Gaps  0;

Qy   892 tggacccaggtgggcaaactcttcctgtaaagaaccagacaggaactatttttaggctct 951
         |||| ||||| ||||| ||||| ||||||||||| |||||| ||||| ||| | |||| |
Db   572 TAGGCCAGGGGTTGGCATACTTTTCTGTAAAGAGCCAGATAGGAAATATTTTCAATTNT 513

Qy   952 gtgtgccatatgtccagtcacaactactcatctctgcctctgtagcacgaaagcaatt 1011
         ||| ||||| |||||  |    ||| ||   ||||||| |||||||| |||||||| ||
Db   512 GTGGGCTTATGGTCCTCCTGTACAACCACTAAATTCTGCAGCTGTAGCATGAAAGCAGCC 453

Qy   1012 agcaacaatatgtcaacaaa 1031
          ||||||| || || ||||
Db    452 ACAGACGTAAGTAAACAAA 433

RESULT 8
AI128823
```

```
LOCUS       AI128823                 393 bp    mRNA    linear   EST 11-SEP-1998
DEFINITION  qa94d06.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1694411 3', mRNA sequence.
ACCESSION   AI128823
VERSION     AI128823.1  GI:3597337
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 393)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..393
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1694411"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT      140 a     70 c     53 g    130 t
ORIGIN
Query Match          3.4%;   Score 77.4;  DB 9;  Length 393;
Best Local Similarity   60.2%;   Pred. No. 5.6e-09;
Matches  174;   Conservative    0;  Mismatches  96;  Indels  19;  Gaps    2;

Qy     901 gttgggcaaactcttcctgtaagaaccagacaggaactatttaggctctgtgtgccat 960
           |||| ||||||| || |||||||| |||||| |  ||||||  ||||  ||| |||||
Db     103 GGTCAGCAAACTTTTCTGTAAAGGGCCATACAATACATATTTCAGTCTCTGCAGACCAT 162

Qy     961 atggtctcagtcacaactactcatctcgcctgtagcacgaaagcaattagcacaat 1020
           ||||| ||||| ||||||||||||  |||| |||||||| |||||| |||||||||
Db     163 TTGGTCTTTGCTGAAACTACTCAACTCTGCTAACATAACACAAAAGCTGCCACAGACAAT 222

Qy    1021 atgtcaacaaacatatgtgaccccatgaaaacttatttat---tatggatacggaaacc 1077
           |||| |||| |||  |||||||| || |||||||||| ||   ||||||||||
Db     223 ATGTAAACAAAATGAGGGTGACTGTAATAAAATGTTATTTATGAACAACTGAAATTTGAACT 282

Qy    1078 tgaaaataatgtcttttct-----------tttgatttttcccaatcattaaaa 1121
           | ||| ||||| ||||||           || |||| |||| ||||||| ||||
Db     283 TCAAATCATTTTCATGCTTCCCAAAACATTCTACTTTTGACTATTTTCAACCATTTAAA 342

Qy    1122 aacgtaaaactactcttaggtcgcaaggttaagccattctcagcttag 1170
           ||| || ||||| || ||||||||||| ||| ||||||| |||| |||
Db     343 AAAGTAAACAAACCATTCTTAGGTGACAGGCTGTAGAAAAACTAGTCATAG 391
BASE COUNT      9
W76494/c    W76494                   411 bp    mRNA    linear   EST 17-OCT-1996
LOCUS
```

```
DEFINITION  zd67d03.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:345701 5', mRNA sequence.
ACCESSION   W76494
VERSION     W76494.1  GI:1386739
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
AUTHORS     Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Putative full length read
            Insert Length: 466    Std Error: 0.00
            Seq primer: mob.REGA+ET.
FEATURES             Location/Qualifiers
     source          1..411
                     /organism="Homo sapiens"
                     /db_xref="GDB:1271076"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:345701"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAGTCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT      129 a     58 c     79 g   145 t
ORIGIN
Query Match          3.4%;   Score 77.4;  DB 10;  Length 411;
Best Local Similarity   60.2%;   Pred. No. 5.6e-09;
Matches  174;   Conservative    0;  Mismatches  96;  Indels  19;  Gaps    2;

Qy     901 gttgggcaaactcttcctgtaagaacagacaggaactatttaggctctgtgtgccat 960
           |||| ||||||| || |||||||| |||||| |  ||||||  ||||  ||| |||||
Db     325 GGTCAGCAAACTTTTCTGTAAAGGGCCATACAATACATATTTCAGTCTCTGCAGACCAT 266

Qy     961 atggtctcagtcacaactactcatctcgcctctgtagcacgaaagcaattagcaacaat 1020
           ||||| ||||| ||||||||||||  |||| |||||||| |||||| |||||||||
Db     265 TTGGTCTTTGCTGAAACTACTCAACTCTGCTAACATAACACAAAAGCTGCCACAGACAAT 206

Qy    1021 atgtcaacaaacatatgtgacccatgaaaacttatttat---tatggatacggaaacc 1077
           |||| |||| |||  |||||||| || |||||||||| ||   ||||||||||
Db     205 ATGTAAACAAAATGAGGGTGACTGTAATAAAATGTTATTTATGAACAACTGAAATTTGAACT 146

Qy    1078 tgaaaataatgtcttttct-----------tttgatttttcccaatcattaaaa 1121
           | ||| ||||| ||||||           || |||| |||| ||||||| ||||
Db     145 TCAAATCATTTTCATGCTTCCCAAAACATTCTACTTTTGACTATTTTCAACCATTTAAA 86
```

```
Qy    1122  aacgtaaaaactactcttaggtcgcaggttaagccattccagcttag 1170
            |||||||||| | |||||||| ||  ||| || |  ||| |||||
Db      85  AAAGTAAAAACCAT"CTTAGGTGACAGCGTGTAGAAAAACTAGTCATAG 37
```

RESULT 10
AG154552
LOCUS   AG154552   680 bp   DNA   linear   GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-018J16.TJ, genomic survey
            sequence.
ACCESSION   AG154552
VERSION   AG154552.1 GI:1684230
KEYWORDS   GSS (genome survey sequence).
SOURCE   Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-018J16.TJ.
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (sites)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   BAC end sequences of Library RPCI-43
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 680)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
            PRIMERS
            Sequencing: TJ
            LIBRARY
            Vector   : pBACe3.6
            R.Site 1 : EcoRI
            R.Site 2 : EcoRI
FEATURES   Location/Qualifiers
    source 1..680
            /organism="Pan troglodytes"
            /db_xref="taxon:9598"
            /clone="RP43-018J16.TJ"
            /sex="male"
            /cell_type="lymphocytes"
            /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT
ORIGIN   207 a   127 c   133 g   213 t

Query Match   3.4%; Score 77.4; DB 12; Length 680;
Best Local Similarity 62.6%; Pred. NO. 6.5e-09;
Matches 137;   Conservative 0; Mismatches 81; Indels 1; Gaps 1;

```
Qy     881  tttccatgctggacccaggtgggcaaactcttcctgtaaagaaccagacaggaacta 940
            |-||||  | ||||||| || |||| |||||| | || |||||||| ||||  ||||
Db     227  TCTCCATATTCTAGAATAGGAGTCAGCAAATTTCTACTACAAAGGGCAAGACAGTAAATA 286

Qy     941  tttaggctctgtgtgccatatggtctccagtcacaacatactactctctgctcctgtagca 1000
            || ||||| ||  |||||| ||||||||||||||||||||| || ||  |||||||||||
Db     287  TTTTAGGCTTTGAGGGCCATACCGTCTCTCGTCATAACTACTACTCAATTCTGCTGTTTTAGCG 346

Qy    1001  cgaaagcaattagcaacaacatatgtcaacaaacatatgtgacccccatgaaaacttattta 1060
            ||||||| ||||||||  ||||  ||    ||   |||   |||||||| ||||| |||||
Db     347  CGAAAGCAGCCATAGACAATACAATACATAAATGAAAAAGCATGGCTGTGTTCCAATAACATTTA 406

Qy    1061  -ttatggatacggaaacctgaaaataatgtcttttcttt 1098
             ||||||||| |||||||||||||||||| || ||||||
Db     407  TTTTACGGACACTGAAATTTGAATTGCATATAATTTTTAT 445
```

RESULT 11
AI926274
LOCUS   AI926274   400 bp   mRNA   linear   EST 02-SEP-1999
DEFINITION wo43g06.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2458138 3',
            mRNA sequence.
ACCESSION   AI926274
VERSION   AI926274.1 GI:5662162
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 400).
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE   Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 382.
FEATURES   Location/Qualifiers
    source 1..400
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2458138"
            /clone_lib="NCI_CGAP_Gas4"
            /tissue_type="poorly differentiated adenocarcinoma with
            signet ring cell features"
            /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.69 kb. Life Technologies catalog #:
            11549-011"
BASE COUNT   133 a   73 c   54 g   140 t
ORIGIN

Query Match   3.4%; Score 77.2; DB 9;   Length 400;
Best Local Similarity 64.9%; Pred. No. 6.3e-09;
Matches 131;   Conservative 0; Mismatches 68; Indels 3; Gaps 1;

```
Qy     901  gtcgggcaaactcctcctgtaaagaaccagacaggaggaactatttaggctctgtgtgccat 960
            ||||||||||||| ||||||| |||||| |||| ||  |  ||| || ||||| || ||||
Db     127  GGTCAGCAAACTTTTCGTAAAGGGCCATACAATACATATTTCAGTCTCTGCAGACCAT 186

Qy     961  atggtccagtcacaactactcatctgcctctgtagcacgaaagcaattagcaacaat 1020
            | ||||| | ||||||| | | |||  |||| ||||||| ||||| ||||||| ||||
Db     187  TTGGTCTTTGCTGAAACTACTCAACTCTGCTAACTAACAAAAAGCTGCCACAGACAAT 246

Qy    1021  atgtcaacaaacatatgtgacccatggaaaactttattatatggatacggaaaacctga 1080
            | ||||| |  || | |||  || |||   |  || | |  ||| ||| || |||||
Db     247  ATGTAAACAAATGAGGGTGACTGACTGTATAAAATGTTAT---TTATGAACACTGAAATTTGA 303

Qy    1081  aaataatgtctttcttttgatt 1102
            |||||| | |||| |||  |||
Db     304  ACTTCAAATCATTTTCATGCTT 325
```

RESULT 12
AI240516/c
LOCUS   AI240516   322 bp   mRNA   linear   EST 30-NOV-1998
DEFINITION qh12h02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1644499 3', mRNA sequence.

```
ACCESSION    AI240516
VERSION      AI240516.1    GI:3835913
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 322)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert Length: 735    Std Error: 0.00
             Seq primer: -40UP from Gibco
             High quality sequence stop: 315.
FEATURES             Location/Qualifiers
     source          1..322
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1844499"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT     90 a     72 c      67 g      93 t
ORIGIN
```

```
Query Match            3.4%;     Score 77;  DB 9;  Length 322;
Best Local Similarity  71.6%;    Pred. No. 6.6e-09;
Matches  101;  Conservative   0;  Mismatches  40;  Indels   0;  Gaps   0;

Qy    891 ctggacccaggttggccaaactcttcctgtaaagaaccaggacaggaactattttaggctc  950
          ||||| ||||||||||||||||||||||||| ||||||| ||||| ||||||||  |||||
Db    178 CTAGACCAGAGGTGGCAAACTCTATCTGTAATTGTCCAGATAGTAAATATTTTAGGTTT  119

Qy    951 tgtgtggccatatgtcagtcacaacaactactcatctctgcctctgtagcacgaaagcaat 1010
          ||||| ||||||||||| ||||||||||||||||||||| |||| ||||| |||| ||
Db    118 TGTGGGCCATAGGGTCTCTATTCCAACTACTCAACTCTGCGTGTTGCAGCATGAAAGTAGC  59

Qy   1011 tagcaacaatatgtcaacaaa 1031
          ||  ||||||||||||
Db     58 CATAGATGTTAAGTCAACAAA  38
```

```
RESULT 13
AA443938
LOCUS        AA443938           386 bp    mRNA    linear    EST  03-JUN-1997
DEFINITION   zv51c03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:757156
             3', mRNA sequence.
ACCESSION    AA443938
VERSION      AA443938.1    GI:2156613
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 386).
AUTHORS      Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
```

```
TITLE        Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
             , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
             ,T., Waterston,R. and Wilson,R.
JOURNAL      WashU-Merck EST Project 1997
             Unpublished (1997)
COMMENT      Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Seq primer: -41m13 fwd. ET from Amersham
             High quality sequence stop: 321.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="GDB:597767"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:757156"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCGCCCAATTTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT    111 a     72 c      68 g     135 t
ORIGIN
```

```
Query Match            3.4%;     Score 77;  DB 9;  Length 386;
Best Local Similarity  63.0%;    Pred. No. 7e-09;
Matches  119;  Conservative   0;  Mismatches  70;  Indels   0;  Gaps   0;

Qy    881 tttcccatgcctggacccaggttggccaaactcttcctgtaaagaaccaggacaggaacta  940
          |||| || ||||| ||| | |||||||| |||||||||||||||| || ||||||||
Db    139 TTTTTTCTATCTAGAACATGAGTGGTTGGCAAACTTTTTCTATAAAGGGCCTGATAATAAATA  198

Qy    941 tttttaggctctgtgtgccatatggtctcagtcacaactactcatctctgcctctgtgca 1000
          ||| ||||| || |||| | ||||||| ||| ||||  |||| ||||  || |||||
Db    199 TTTCAGGCTTTGAGGGTCACATGGTCTCTGTAGCACCTACTCAATTCTGCCATTATAGTA  258

Qy   1001 cgaaagcaattagcaacaactactcaacaacatatgtgaccccatgaaaacttttattta 1060
          | |||      ||||||||||  |||||||| | ||  |||||  |  ||     |   |
Db    259 CAGAAGCTGCCACAGACCAATACGTCAACAAAATAAGAGTGGCTGTGTGCCAATAAAACTTTT  318

Qy   1061 ttatggata 1069
           |||||||
Db    319 ATTTCCAAA  327
```

```
RESULT 14
AA444117/c
LOCUS        AA444117           393 bp    mRNA    linear    EST  03-JUN-1997
DEFINITION   zv51c03.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:757156
             5', mRNA sequence.
ACCESSION    AA444117
VERSION      AA444117.1    GI:2156792
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 393)
AUTHORS      Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
```

Left column

AUTHORS   Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
,T., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
,T., Waterston,R. and Wilson,R.
TITLE   WashU-Merck EST Project 1997
JOURNAL   Unpublished (1997)
COMMENT   Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 382.
FEATURES   Location/Qualifiers
source   1..393
/organism="Homo sapiens"
/db_xref="GDB:5977767"
/db_xref="taxon:9606"
/clone="IMAGE:757156"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   120 a   80 c   80 g   113 t
ORIGIN

Query Match   3.4%;   Score 77;   DB 9;   Length 393;
Best Local Similarity   63.0%;   Pred. No. 7e-09;
Matches 119;   Conservative   0;   Mismatches 70;   Indels   0;   Gaps   0;

Qy   881 tttcccatgcctgacccaggttgggcaaactcctgtaaagaaccagacaggaacta 940
          ||| || || | || ||| ||||||| || ||||| ||||| |||| | | ||
Db   380 TTTTTTCTATCTAGAACATGGTTGGCAAACTTTTTTCTATAAACGGCCTGATAATAAATA 321

Qy   941 tttaggctctgtgtgccatatggtctcagtcacaactactcatctctgcctctgtagca 1000
          ||| |||| | | |||||| |||||||| ||| || ||||||| | ||| ||| ||||
Db   320 TTTCAGGCTTTGAGGGTCACATGGTCTCTGTAGCACCTACTCACTTCTGCCATTATAGTA 261

Qy   1001 cgaaagcaattagcaacaatatgtcaacaaacatatgtgacccatgaaaacttatttta 1060
          | ||||| ||| | || ||| | ||||||| |||| || ||| | |||| || || |
Db   260 CAGAAGCTGCCACAGACAATACGTCAACAAATAAGAGTGGCTGTGTGCCAATAAAACTTT 201

Qy   1061 ttatggata 1069
          | || | ||
Db   200 ATTTCCAAA 192

RESULT 15
AI208768
LOCUS   AI208768   417 bp   mRNA   linear   EST 29-NOV-1998
DEFINITION   qg38c10.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1837458
3', mRNA sequence.
ACCESSION   AI208768
VERSION   AI208768.1 GI:3770710
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 417)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Right column

TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 605   Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 412.
FEATURES   Location/Qualifiers
source   1..417
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1837458"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   126 a   78 c   75 g   138 t
ORIGIN

Query Match   3.4%;   Score 77;   DB 9;   Length 417;
Best Local Similarity   63.0%;   Pred. No. 7.2e-09;
Matches 119;   Conservative   0;   Mismatches 70;   Indels   0;   Gaps   0;

Qy   881 tttcccatgcctgacccaggttgggcaaactcctgtaaagaaccagacaggaacta 940
          ||| || || | || ||| ||||||| || ||||| ||||| |||| | | ||
Db   138 TTTTTTCTATCTAGAACATGGTTGGCAAACTTTTTTCTATAAACGGCCTGATAATAAATA 197

Qy   941 tttaggctctgtgtgccatatggtctcagtcacaactactcatctctgcctctgtagca 1000
          ||| |||| | | |||||| |||||||| ||| || ||||||| | ||| ||| ||||
Db   198 TTTCAGGCTTTGAGGGTCACATGGTCTCTGTAGCACCTACTCACTTCTGCCATTATAGTA 257

Qy   1001 cgaaagcaattagcaacaatatgtcaacaaacatatgtgacccatgaaaacttatttta 1060
          | ||||| ||| | || ||| | ||||||| |||| || ||| | |||| || || |
Db   258 CAGAAGCTGCCACAGACAATACGTCAACAAATAAGAGTGGCTGTGTGCCAATAAAACTTT 317

Qy   1061 ttatggata 1069
          | || | ||
Db   318 ATTTCCAAA 326

Search completed: June 6, 2002, 10:43:34
Job time: 8926 sec

us-09-515-369b-1.rst

```
          GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    June  6, 2002, 08:39:08 ; Search time 93.24 Seconds
                                     (without alignments)
                                     6022.289 Million cell updates/sec

Title:          US-09-515-369B-1
Perfect score:  2286
Sequence:       1 taatacgactcactatagg.........tgacttccacggctgggacg 2286

Scoring table:  IDENTITY.NUC
                Gapop 10.0 , Gapext 1.0

Searched:       383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters:    767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_NA:*
            1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	71.8	3.1	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	67.6	3.0	6235	4	US-09-305-384-5	Sequence 5, Appl
	3	67.6	3.0	6679	4	US-09-305-384-1	Sequence 1, Appli
	4	62.2	2.7	178	4	US-09-991-789A-210	Sequence 210, App
	5	62.2	2.7	178	4	US-09-062-451-210	Sequence 210, App
c	6	52.6	2.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	7	52	2.3	3786	4	US-08-975-762-42	Sequence 42, Appl
c	8	52	2.3	3786	4	US-09-295-028-42	Sequence 42, Appl
	9	50.6	2.2	3786	4	US-09-106-582-42	Sequence 42, Appl
	10	50.6	2.2	502	2	US-08-967-101-91	Sequence 91, Appl
	11	50.6	2.2	502	2	US-08-592-541-91	Sequence 91, Appl
	12	50.6	2.2	502	3	US-09-124-698-91	Sequence 91, Appl
	13	50.6	2.2	502	4	US-08-496-841C-91	Sequence 91, Appl
	14	50.6	2.2	502	4	US-08-967-101-10	Sequence 10, Appl
	15	50.6	2.2	1883	2	US-08-967-101-153	Sequence 153, Appl
	16	50.6	2.2	1883	2	US-08-592-541-10	Sequence 10, Appl
	17	50.6	2.2	1883	2	US-08-592-541-153	Sequence 153, App
	18	50.6	2.2	1883	3	US-08-888-077A-8	Sequence 8, Appli
	19	50.6	2.2	1883	3	US-09-124-698-10	Sequence 10, Appl
	20	50.6	2.2	1883	3	US-09-124-698-153	Sequence 153, App
	21	50.6	2.2	1883	4	US-09-127-480-10	Sequence 10, Appl
c	22	50.6	2.2	1883	4	US-09-127-480-153	Sequence 153, App
	23	50.6	2.2	1883	4	US-08-496-841C-10	Sequence 10, Appl
c	24	50.6	2.2	1883	4	US-08-496-841C-153	Sequence 153, App
c	25	48.2	2.1	80246	4	US-09-078-294-4	Sequence 4, Appli
c	26	48.2	2.1	80595	4	US-09-078-294-3	Sequence 3, Appli
c	28	45.8	2.0	1734	2	US-08-858-052-2	Sequence 2, Appli
c	29	45.8	2.0	1734	3	US-09-200-284-2	Sequence 2, Appli
c	30	43.6	1.9	460	2	US-08-967-101-48	Sequence 48, Appl
c	31	43.6	1.9	460	3	US-08-592-541-48	Sequence 48, Appl
c	32	43.6	1.9	460	3	US-09-124-698-48	Sequence 48, Appl
c	33	43.6	1.9	460	4	US-09-127-480-48	Sequence 48, Appl
c	34	42.8	1.9	59065	4	US-08-496-841C-48	Sequence 48, Appl
c	35	42.8	1.9	59065	4	US-09-813-817-3	Sequence 3, Appli
	36	42.4	1.9	289	4	US-09-007-005-17	Sequence 17, Appl
c	37	42.4	1.9	289	4	US-09-244-796-17	Sequence 17, Appl
	38	41.4	1.8	703	4	US-09-313-300-6	Sequence 6, Appli
	39	41.4	1.8	176373	4	US-09-128-155-17	Sequence 17, Appl
c	40	41.2	1.8	7505	4	US-09-078-294-13	Sequence 13, Appl
c	41	39.8	1.7	350	2	US-08-332-766A-20	Sequence 20, Appl
	42	39.2	1.7	84495	4	US-09-797-906-3	Sequence 3, Appli
	43	39.2	1.7	152331	3	US-09-128-155-16	Sequence 16, Appl
	44	38.6	1.7	43795	3	US-08-742-185-101	Sequence 101, App
	45	38.2	1.7	1732	1	US-08-480-346-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08233463
; Patent No. 5670367
; GENERAL INFORMATION:
;  APPLICANT: DORNER, F.
;  APPLICANT: SCHEIFLINGER, F.
;  APPLICANT: FALKNER, F. G.
;  TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;  NUMBER OF SEQUENCES: 52
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 1800 Diagonal Road, Suite 500
;   CITY: Alexandria
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22313-0299
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/232,463
;   FILING DATE:
;   CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/935,313
;   FILING DATE:
;   APPLICATION NUMBER: EP 91 114 300.6
;   FILING DATE: 26-AUG-1991
;  ATTORNEY/AGENT INFORMATION:
;   NAME: BENT, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)836-9300
;   TELEFAX: (703)683-4109
;   TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 7218 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;  IMMEDIATE SOURCE:
;   CLONE: pTZgpt-Fls
US-08-232-463-14
```

Query Match   3.1%;   Score 71.8; DB 1; Length 7218;
Best Local Similarity   4.0%; Pred. No. 9.7e-12;
Matches 16; Conservative 237; Mismatches 144; Indels 0; Gaps 0;

Qy   159 agctcatcggaggctgaggcccaggcacatgtttgcctgaactatccatgttatatgatt 218
          ||||| ||||||||| ||||| ::::::::::::YYYYYYYYYYYYYYYYYYYYYY ::
Db   1047 AGGTCGAGGAGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1106

Qy   219 cctcctcagacagagtgagctactcacgatccagvgtgtaccctgaggcagccaggt 278
          :::::::::: :: ::: ::: YYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db   1107 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1166

Qy   279 gtatccatgacctgcctctgttccagctgcccttaacagctcatcccacctgcct 338
          : ::::::::: ::::: ::::: ::: :::
Db   1167 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1226

Qy   339 gccctcccgccatctgcagacagtagtcctaggattcagctgcctggggctcattt 398
          ::::::::::::::: ::: YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db   1227 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1286

Qy   399 tccctctcagcttcctgctttagctgtctctcctgcctccactattactccagcac 458
          ::::::::: :::::: ::: YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db   1287 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1346

Qy   459 tctcaactggtctctcttttctgtctcatcactgcctcttgacatctttatctcatagtag 518
          ::::::::: ::::::::: YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY :; :
Db   1347 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1406

Qy   519 ttagttaggggttcttggtaatgccctaaatccacat 555
          :: :|| ||
Db   1407 YYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAAT 1443

RESULT 2
US-09-305-384-5
; Sequence 5, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard F
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,649
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 6235;
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-384-5

Query Match   3.0%;   Score 67.6; DB 4; Length 6235;
Best Local Similarity   63.7%; Pred. No. 1.8e-10;
Matches 135; Conservative 0; Mismatches 74; Indels 3; Gaps 2;

Qy   921 aagaaccagacaggaactcatttaggctctgtgtgccatatggtctccagtcacaactac 980
          ||||||||| |||||||| ||||||||||| || ||||||||||||| ||||||
Db   622 aaagatccagatagtcaattttttaggct-tgtggggccgtatggtcctgtcacaatcac 680

Qy   981 tcatctctgcctcgtagcacggaaagcaattagcaacaatatgtcaacaaacatatgtga 1040
          || | ||| || |||||||| |||||| || |||||| || | | || |
Db   681 tctgccctgtcttctctgcacacaaagcagctataaacagctataaaacatacacatgaattttttata 740

Qy   1041 cccatgaaaacttattcttatatggatacggaaacctgaaaataataatgtcttctctttga 1100
          ||||| |||||| ||| || |||||| || || |||| | || |
Db   741 gacatcgagatttgaattcatgatt--ttacatttatcaaaaataatctttttaaaa 798

Qy   1101 ttttttcccaatcattaaaaaacgtaaaac 1132

RESULT 3
US-09-305-384-1
; Sequence 1. Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard F
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,649
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-384-1

Query Match   3.0%;   Score 67.6; DB 4; Length 6679;
Best Local Similarity   63.7%; Pred. No. 1.9e-10;
Matches 135; Conservative 0; Mismatches 74; Indels 3; Gaps 2;

Qy   921 aagaaccagacaggaactcatttaggctctgtgtgccatatggtctccagtcacaactac 980
          ||||||||| |||||||| ||||||||||| || ||||||||||||| ||||||
Db   641 aaagatccagatagtcaattttttaggct-tgtggggccgtatggtcctgtcacaatcac 699

Qy   981 tcatctctgcctcgtagcacggaaagcaattagcaacaatatgtcaacaaacatatgtga 1040
          || | ||| || |||||||| |||||| || |||||| || | | || |
Db   700 tctgccctgtcttctctgcacacaaagcagctataaacagctataaaacatacacatgaattttttata 759

Qy   1041 cccatgaaaacttattcttatatggatacggaaacctgaaaataataatgtcttctcttga 1100
          ||||| |||||| ||| || |||||| || || |||| | || |
Db   760 gacatcgagatttgaattcatgatt--ttacatttatcaaaaataatctttttaaaa 817

Qy   1101 ttttttcccaatcattaaaaaacgtaaaac 1132
          ||||| ||| || ||||||||| ||||||
Db   818 atttccctaaccattaaaagtgaaaagc 849

RESULT 4
US-08-991-789A-210
; Sequence 210, Application US/08991789
; Patent No. 6225054
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;   TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Seed IP Law Group
;   STREET: 701 Fifth Avenue, Suite 6300
;   CITY: Seattle
;   STATE: Washington
;   COUNTRY: USA
;   ZIP: 98104-7092
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/991,789A

Db   799 atttccctaaccattaaaagtgaaaagc 830

; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 210:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 210:
US-08-991-789A-210

Query Match   2.7%; Score 62.2; DB 4; Length 178;
Best Local Similarity 71.3%; Pred. No. 9.6e-10;
Matches   82; Conservative   0; Mismatches   33; Indels   0; Gaps   0;

Qy   891 ctggacccaggttggcaaactccttcctgtaaagaaccagacaggaactatttaggctc 950
           |||| ||||||| ||| |||| || || |||||||||||| |||||| |||| ||||||
Db   20 CTAGAGTAGGGTTTGGCCAACTTTTTCTATAAAGGACCAGAGAGTAAATATTTCAGGCTT 79

Qy   951 tgtgtgccatatggctcagtcacaactactcatctctgcctctgtagcacgaa 1005
           |||| | || | |||||| ||||||||| |||||||| |||||| ||
Db   80 TGTGGTTGTGCAGTCTCTTGCAACTACTCAGCTCTCCCATTGCCATGAGCATAGAA 134

RESULT 5
US-09-062-451-210
; Sequence 210 Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 210:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-062-451-210

Query Match   2.7%; Score 62.2; DB 4; Length 178;
Best Local Similarity 71.3%; Pred. No. 9.6e-10;
Matches   82; Conservative   0; Mismatches   33; Indels   0; Gaps   0;

Qy   891 ctggacccaggttggcaaactccttcctgtaaagaaccagacaggaactatttaggctc 950
           |||| ||||||| ||| |||| || || |||||||||||| |||||| |||| ||||||
Db   20 CTAGAGTAGGGTTTGGCCAACTTTTTCTATAAAGGACCAGAGAGTAAATATTTCAGGCTT 79

Qy   951 tgtgtgccatatggctcagtcacaactactcatctctgcctctgtagcacgaa 1005
           |||| | || | |||||| ||||||||| |||||||| |||||| ||
Db   80 TGTGGTTGTGCAGTCTCTTGCAACTACTCAGCTCTCCCATTGCCATGAGCATAGAA 134

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match   2.3%; Score 52.6; DB 1; Length 7218;
Best Local Similarity   9.0%; Pred. No. 9.4e-06;
Matches   28; Conservative   162; Mismatches   121; Indels   0; Gaps   0;

Qy   484 catcactgcctcttgacatcttatctcatagtagtagttagtagggtccttggtaatgcc 543
          ||||||| | |||||| | || | | | ||| ||||
Db   1490 CATCACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGAATAGAAGAAGAATTTGGTACRRRR 1431

```
Qy   544 ctaaatccacatgtgggaagggggagtggggggaaagagagtgcgctgtgggggctgtgcc 603
         ::: ::: : :::::::::::::: ::::::::::::::
Db  1430 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1371

Qy   604 tacttctggagggtaagactcgggccctccaggaacaaaggattcaggctggtggcagct 663
         :::  ::: :::  ::
Db  1370 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1311

Qy   664 atagccaagcagactgctggcccaggattgcaaaggagtgatattttgttgcttaagaaat 723
         : :: ::: ::
Db  1310 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1251

Qy   724 aaacaacactgagtatgagatgagatgagagggggtgttggtgccagagagattgggaagag 783
         ::: :: : :: :::::::: :::::::
Db  1250 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1191

Qy   784 tctgccaaggg 794
Db  1190 RRRRRRRRRRR 1180


RESULT 7
US-08-975-762-42
; Sequence 42, Application US/08975762
; Patent No. 6207169
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G.
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Houghton, Raymond
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND   TREATMEN
;   NUMBER OF SEQUENCES: 73
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  SEED and BERRY LLP
;     STREET: 6300 Columbia Center, 701 Fifth Avenue
;     CITY: Seattle
;     STATE: Washington
;     COUNTRY: USA
;     ZIP: 98104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/975,762
;     FILING DATE: 21-MAR-1997
;     CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Maki, David J.
;     REGISTRATION NUMBER: 31,392
;     REFERENCE/DOCKET NUMBER: 210121.439
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 206-622-4900
;     TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 42:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 3786 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-975-762-42

  Query Match             2.3%;  Score 52;  DB 4;  Length 3786;
  Best Local Similarity   67.2%;  Pred. No. 9.7e-06;
  Matches   90;  Conservative    0;  Mismatches   40;  Indels   4;  Gaps   1;

Qy   914 tcctgtaaagaaccagacaggaactattttagctcgtgtgccatatggtc---tca 969
         || |||| | |||||||| | |||||| || |||| |||||||||| ||||||
Db  3296 TTTCTGTACTGCCAGACAGTAAACATTTGGACTTTGTATATGGTCTCTTTCT 3355

Qy   970 gtcacaactactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaaca 1029
```

```
Db  3356 GTTGCAACTACTGAACTCTTCCATTGTAGCACGAAGCGGCTGCGACGAAGCGGCGTCGCAGAACAATATGTAAACA 3415
         ||  ||||||||| ||||||||||||| || |||||||||||||||||||||||||||||||||||||||||||| ||||

Qy  1030 aacatatgtgaccc 1043
         | |||| |
Db  3416 GATGAGCATGACTC 3429


RESULT 8
US-09-295-028-42
; Sequence 42, Application US/09295028
; Patent No. 6277381
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G.
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Houghton, Raymond L.
;   APPLICANT: McNeill, Patricia D.
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;   TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
;   FILE REFERENCE: 210121.439C4
;   CURRENT APPLICATION NUMBER: US/09/295,028
;   CURRENT FILING DATE: 1999-04-20
;   NUMBER OF SEQ ID NOS: 85
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 42
;   LENGTH: 3786
;   TYPE: DNA
;   ORGANISM: Ehrlichia sp.
US-09-295-028-42

  Query Match             2.3%;  Score 52;  DB 4;  Length 3786;
  Best Local Similarity   67.2%;  Pred. No. 9.7e-06;
  Matches   90;  Conservative    0;  Mismatches   40;  Indels   4;  Gaps   1;

Qy   914 tccctgtaaagaaccagacaggaactattttagctcgtgtgccatatggtc---tca 969
         || ||| | |||||||| | |||||| || |||| |||||||||| ||||||
Db  3296 ttctgtactgccagacagtaaacatttggactttgtatatggtctctttct 3355

Qy   970 gtcacaactactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaaca 1029
         | |||| | |||||| | ||||| ||||||| |||||||| || ||||||||||||
Db  3356 gttgcaactactgaactcttccattgtagcacgaagcgggctgcgacgaagcggctgaacatatgtaaaca 3415

Qy  1030 aacatatgtgacctc 1043
         | |||| |
Db  3416 gatgagcatgactc 3429


RESULT 9
US-09-106-582-42
; Sequence 42, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G.
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Houghton, Raymond
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND   THERAP
;   NUMBER OF SEQUENCES: 73
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  SEED and BERRY LLP
;     STREET: 6300 Columbia Center, 701 Fifth Avenue
;     CITY: Seattle
;     STATE: Washington
;     COUNTRY: USA
;     ZIP: 98104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/106,582
;     FILING DATE: 29-JUN-1998
```

```
;       CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Maki, David J.
;       REGISTRATION NUMBER: 31,392
;       REFERENCE/DOCKET NUMBER: 210121.439C2
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 206-622-4900
;       TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 42:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 3786 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
US-09-106-582-42

Query Match            2.3%;    Score 52;  DB 4;  Length 3786;
Best Local Similarity 67.2%;    Pred. No. 9.7e-06;
Matches   90;  Conservative  0;  Mismatches  40;  Indels  4;  Gaps  1;

Qy   914 ttcctgtaaagaaccagacaggaactatttaggctctgtgtgccatatggtc---tca 969
         || |||| |||||||||| |||| | |||||  ||||| |||  ||||||||     ||
Db  3296 TTTCTGTGTACTGCCAGACAGTAAACATTTTGGACTTTGTATGTCTCTTTCT       3355

Qy   970 gtcacaactactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaaca 1029
         || ||||||||| | |||||||| | |||||||||||||| ||| ) |||||||| |||
Db  3356 GTTGCAACTACTGAACTCTTCCATTGTAGCACGAAGGCGGCTGCGACAAATATGTAAACA 3415

Qy  1030 aacatatgtgaccc 1043
         | ||| |
Db  3416 GATGAGCATGACTC 3429


RESULT 10
US-08-967-101-91
; Sequence 91, Application US/08967101
; Patent No. 5840540
;   GENERAL INFORMATION:
;       APPLICANT: ST. GEORGE-HYSLOP, PETER H
;       APPLICANT: ROMMENS, JOHANNA M
;       APPLICANT: FRASER, PAUL E
;       TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;       TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;       NUMBER OF SEQUENCES: 183
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;           STREET: High Street Tower - 125 High Street
;           CITY: Boston
;           STATE: Massachusetts
;           COUNTRY: U.S.A.
;           ZIP: 02110
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/08/967,101
;           FILING DATE: 10-NOV-1997
;           CLASSIFICATION: 435
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 08/592,541
;           FILING DATE:
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Pitcher, Edmund R.
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (617) 248-7000
;           TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 91:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 502 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
US-08-967-101-91

Query Match            2.2%;    Score 50.6;  DB 2;  Length 502;
Best Local Similarity 64.3%;    Pred. No. 7.6e-06;
Matches   92;  Conservative  0;  Mismatches  49;  Indels  2;  Gaps  1;

Qy   917 ctgtaaagaaccagacaggaactatttaggctctgtgtgccatatggtctcagtcacaa 976
         |||||||| |||||||| | |||||||| |  |||||| ||||||||||||||||||||
Db   264 CTGTAAAAGCCAGACTG--AATATTTAAGCTCTCATATGGTCTCCAGGGCAA 321

Qy   977 ctactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatat 1036
         ||||| ||| ||||| ||||||||||| ||| |||||||||||||| | | |||||| ||
Db   322 ACACTCAACTGTGTCTACTGTAGTGTGAAAGCAGGCACCAGACAATGTATTAACCAAGGAGG 381

Qy  1037 gtgacccatgaaaactttattt 1059
         ||| || |||  || || | |||
Db   382 GTGGTCACTTTCCAATGAAAGTT 404


RESULT 11
US-08-592-541-91
; Sequence 91, Application US/08592541
; Patent No. 5986054
;   GENERAL INFORMATION:
;       APPLICANT: ST. GEORGE-HYSLOP, PETER H
;       APPLICANT: ROMMENS, JOHANNA M
;       APPLICANT: FRASER, PAUL E
;       TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;       TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;       NUMBER OF SEQUENCES: 183
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;           STREET: High Street Tower - 125 High Street
;           CITY: Boston
;           STATE: Massachusetts
;           COUNTRY: U.S.A.
;           ZIP: 02110
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/08/592,541
;           FILING DATE:
;           CLASSIFICATION: 800
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Pitcher, Edmund R.
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (617) 248-7000
;           TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 91:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 502 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
US-08-592-541-91

Query Match            2.2%;    Score 50.6;  DB 2;  Length 502;
Best Local Similarity 64.3%;    Pred. No. 7.6e-06;
Matches   92;  Conservative  0;  Mismatches  49;  Indels  2;  Gaps  1;

Qy   917 ctgtaaagaaccagacaggaactatttaggctctgtgtgccatatggtctcagtcacaa 976
         |||||||| |||||||| | |||||||| |  |||||| ||||||||||||||||||||
Db   264 CTGTAAAAGCCAGACTG--AATATTTAAGCTCTCATATGGTCTCCAGGGCAA 321
```

Left column

```
Qy   977  ctactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatat 1036
              |||||| || ||| ||| |||||   |||| | ||| ||
Db   322  ACACTCAACTGTGCTACTGTAGTGTGAAAGCAGGCACAGACAATGTATTAACCAAGGAGG 381

Qy  1037  gtgacccatgaaaacttattt 1059
              ||| | || |  ||| ||
Db   382  GTGGTCACTTTCCAATGAAAGTT 404
```

RESULT 12
US-09-124-698-91
; Sequence 91, Application US/09124698
; Patent No. 6117978
; GENERAL INFORMATION:
;   APPLICANT: ST. GEORGE-HYSLOP, PETER H
;   APPLICANT: ROMMENS, JOHANNA M
;   APPLICANT: FRASER, PAUL E
;   TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;   TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;   NUMBER OF SEQUENCES: 183
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;   STREET: High Street Tower - 125 High Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/124,698
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/592,541
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Pitcher, Edmund R.
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 91:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 502 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-09-124-698-91

Query Match   2.2%;   Score 50.6;   DB 3;   Length 502;
Best Local Similarity 64.3%;   Pred. NO. 7.6e-06;
Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

```
Qy   917  ctgtaaagaaccagacaggaactatttaggctctgtgtgccatatggtctcagtcacaa 976
              |||||| | |||| ||| ||  ||||||| | |||||||| ||||||
Db   264  CTGTAAAAGCCAGACTG--AATATTTAAGCTCTATGGGTCATATGGTCTCCAGGGCAA 321

Qy   977  ctactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatat 1036
              |||||| || ||| ||| |||||   |||| | ||| ||
Db   322  ACACTCAACTGTGCTACTGTAGTGTGAAAGCAGGCACAGACAATGTATTAACCAAGGAGG 381

Qy  1037  gtgacccatgaaaacttattt 1059
              ||| | || |  ||| ||
Db   382  GTGGTCACTTTCCAATGAAAGTT 404
```

RESULT 13

Right column

US-09-127-480-91
; Sequence 91, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
;   APPLICANT: ST. GEORGE-HYSLOP, PETER H
;   APPLICANT: ROMMENS, JOHANNA M
;   APPLICANT: FRASER, PAUL E
;   TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;   TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;   NUMBER OF SEQUENCES: 183
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;   STREET: High Street Tower - 125 High Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/127,480
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/592,541
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Pitcher, Edmund R.
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 91:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 502 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-09-127-480-91

Query Match   2.2%;   Score 50.6;   DB 4;   Length 502;
Best Local Similarity 64.3%;   Pred. NO. 7.6e-06;
Matches 92; Conservative 0; Mismatches 49; Indels 2; Gaps 1;

```
Qy   917  ctgtaaagaaccagacaggaactatttaggctctgtgtgccatatggtctcagtcacaa 976
              |||||| | |||| ||| ||  ||||||| | |||||||| ||||||
Db   264  CTGTAAAAGCCAGACTG--AATATTTAAGCTCTATGGGTCATATGGTCTCCAGGGCAA 321

Qy   977  ctactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatat 1036
              |||||| || ||| ||| |||||   |||| | ||| ||
Db   322  ACACTCAACTGTGCTACTGTAGTGTGAAAGCAGGCACAGACAATGTATTAACCAAGGAGG 381

Qy  1037  gtgacccatgaaaacttattt 1059
              ||| | || |  ||| ||
Db   382  GTGGTCACTTTCCAATGAAAGTT 404
```

RESULT 14
US-08-496-841C-91
; Sequence 91, Application US/08496841C
; Patent No. 6210919
; GENERAL INFORMATION:
;   APPLICANT: ST. GEORGE-HYSLOP, PETER H
;   APPLICANT: ROMMENS, JOHANNA M
;   APPLICANT: FRASER, PAUL E
;   TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;   TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;   NUMBER OF SEQUENCES: 175
;   CORRESPONDENCE ADDRESS:

```
                                            ADDRESSEE: Darby & Darby, PC
                                            STREET: 805 Third Avenue
                                            CITY: New York
                                            STATE: New York
                                            COUNTRY: U.S.A.
                                            ZIP: 10022
                                    COMPUTER READABLE FORM:
                                            MEDIUM TYPE: Floppy disk
                                            COMPUTER: IBM PC compatible
                                            OPERATING SYSTEM: PC-DOS/MS-DOS
                                            SOFTWARE: PatentIn Release #1.0, Version #1.30
                                    CURRENT APPLICATION DATA:
                                            APPLICATION NUMBER: US/08/496,841C
                                            FILING DATE: 28-Jun-1995
                                            CLASSIFICATION: <unknown>
                                    ATTORNEY/AGENT INFORMATION:
                                            NAME: Paul F. Fehlner, Ph.D.
                                            REGISTRATION NUMBER: 35,135
                                    TELECOMMUNICATION INFORMATION:
                                            TELEPHONE: (212) 527-7700
                                            TELEFAX: (212) 753-6237
                            INFORMATION FOR SEQ ID NO: 91:
                                    SEQUENCE CHARACTERISTICS:
                                            LENGTH: 502 base pairs
                                            TYPE: nucleic acid
                                            STRANDEDNESS: single
                                            TOPOLOGY: linear
                                            MOLECULE TYPE: DNA (genomic)
                                    SEQUENCE DESCRIPTION: SEQ ID NO: 91:
                            US-08-496-841C-91

                            Query Match          2.2%;   Score 50.6;  DB 4;  Length 502;
                            Best Local Similarity  64.3%;   Pred. No. 7.6e-06;  Indels  2;  Gaps  1;
                            Matches  92;  Conservative  0;  Mismatches  49;

                            Qy    917 ctgtaaagaaccagacaggaactattttaggctctgtgtgccatatggtctcagtcacaa 976
                                      ||||||||||| ||||||| | |||||| |||||| || | |||||||||||||||||
                            Db    264 CTGTAAAAAGCCAGACTG--AATATTTTAAGCTCTATGGGTCATATGGTCTCCAGGGCAA 321

                            Qy    977 ctactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatat 1036
                                      ||||| |||| ||| ||||||  |||||| |||||| || || |  |||||||||
                            Db    322 ACACTCAACTGTGCTACTGTAGTGTGAAAGCAGGCACACGACAATGTATTAACCAAGGAGG 381

                            Qy   1037 gtgacccatgaaaactttattt 1059
                                      ||| || | ||
                            Db    382 GTGGTCACTTTCCAATGAAAGTT 404


                    RESULT 15
                    US-08-967-101-10/c
                    ; Sequence 10, Application US/08967101
                    ; Patent No. 5840540
                    ; GENERAL INFORMATION:
                    ; APPLICANT: ST. GEORGE-HYSLOP, PETER H
                    ; APPLICANT: ROMMENS, JOHANNA M
                    ; APPLICANT: FRASER, PAUL E
                    ; TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
                    ; TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
                    ; NUMBER OF SEQUENCES: 183
                    ; CORRESPONDENCE ADDRESS:
                    ;   ADDRESSEE: TESTA, HURWITZ & THIBEAULT
                    ;   STREET: High Street Tower - 125 High Street
                    ;   CITY: Boston
                    ;   STATE: Massachusetts
                    ;   COUNTRY: U.S.A.
                    ;   ZIP: 02110
                    ; COMPUTER READABLE FORM:
                    ;   MEDIUM TYPE: Floppy disk
                    ;   COMPUTER: IBM PC compatible
                    ;   OPERATING SYSTEM: PC-DOS/MS-DOS
                    ;   SOFTWARE: PatentIn Release #1.0, Version #1.30
                    ; CURRENT APPLICATION DATA:
```

```
            APPLICATION NUMBER: US/08/967,101
            FILING DATE: 10-NOV-1997
            CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
            APPLICATION NUMBER: 08/592,541
            FILING DATE:
    ATTORNEY/AGENT INFORMATION:
            NAME: Pitcher, Edmund R.
    TELECOMMUNICATION INFORMATION:
            TELEPHONE: (617) 248-7000
            TELEFAX: (617) 248-7100
    INFORMATION FOR SEQ ID NO: 10:
            SEQUENCE CHARACTERISTICS:
            LENGTH: 1883 base pairs
            TYPE: nucleic acid
            STRANDEDNESS: single
            TOPOLOGY: linear
            MOLECULE TYPE: cDNA
    US-08-967-101-10

    Query Match          2.2%;   Score 50.6;  DB 2;  Length 1883;
    Best Local Similarity  64.3%;   Pred. No. 1.7e-05;  Indels  2;  Gaps  1;
    Matches  92;  Conservative  0;  Mismatches  49;

    Qy    917 ctgtaaagaaccagacaggaactattttaggctctgtgtgccatatggtctcagtcacaa 976
              ||||||||||| ||||||| | |||||| |||||| || | |||||||||||||||||
    Db    239 CTGTAAAAAGCCAGACTG--AATATTTTAAGCTCTATGGGTCATATGGTCTCCAGGGCAA 182

    Qy    977 ctactcatctctgcctctgtagcacgaaagcaattagcaacaatatgtcaacaaacatat 1036
              ||||| |||| ||| ||||||  |||||| |||||| || || |  |||||||||
    Db    181 ACACTCAACTGTGCTACTGTAGTGTGAAAGCAGGCACACGACAATGTATTAACCAAGGAGG 122

    Qy   1037 gtgacccatgaaaactttattt 1059
              ||| || | ||
    Db    121 GTGGTCACTTTCCAATGAAAGTT 99


    Search completed: June  6, 2002, 11:55:02
    Job time: 11754 sec
```

Thu Jun 6 16:21:05 2002

us-09-515-369b-1.rni